

## Implementasi Data Mining Untuk Pengelompokan Santri Menggunakan Metode K-Medoids

Syamsri Hasrianti<sup>1\*</sup>, Evi Dianti Bintari<sup>2</sup>, Cindy Pangestika<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara  
Email: <sup>1\*</sup>syamsriaci@gmail.com, <sup>2</sup>evidanti@ppkia.ac.id, <sup>3</sup>cindy@ppkia.ac.id

### Abstrak

Clustering merupakan suatu proses pengelompokan data, observasi atau mengelompokkan kelas yang memiliki kesamaan objek. Pondok pesantren daarul ilmi boarding school juga harus melakukan pengelompokan atau clustering terhadap santri santrinya untuk memaksimalkan proses pengajaran. Pondok pesantren tersebut belum memiliki sebuah aplikasi komputer untuk pengelompokan santri sehingga sering terjadi kesalahan dalam menempatkan santri sesuai kriteria yang telah ditetapkan oleh pihak pesantren. Proses pengelompokan santri dilakukan berdasarkan nilai ahlak, ibadah, rata-rata nilai raport, prestasi dan nilai tahfidz. Proses pemberian nilai yang dilakukan masih bersifat subjektif, artinya pemberian nilai didasarkan atas pemberian nilai dari guru pengampu mata pelajaran tertentu, sehingga penelitian ini bertujuan untuk membantu menentukan keputusan pengelompokan santri secara objektif. Penelitian ini menggunakan Algoritma K-Medoids atau yang biasa disebut dengan Partitioning around method (PAM). Penelitian ini menggunakan 54 data santri yang aktif pada tahun ajaran 2021/2022 dimana santri akan dikelompokkan dalam kelas murtafi dan mumtaz. Penelitian ini meliputi tahap penentuan jumlah cluster, pemilihan medoid awal secara acak, menghitung jarak masing-masing objek menggunakan manhattan dan minkowski distance serta menghitung total simpangan. Berdasarkan uji coba program yang dilakukan, penelitian ini berhasil mengelompokkan santri ke dalam kelas murtafi dan mumtaz. Berdasarkan perhitungan jarak menggunakan manhattan, 15 santri masuk dalam kelas murtafi dan 39 santri masuk dalam kelas mumtaz, sedangkan jika menggunakan perhitungan jarak minkowski, 14 santri masuk dalam kelas murtafi dan 40 santri masuk dalam kelas mumtaz. Dari evaluasi menggunakan Davies Bouldin Index (DBI) terhadap hasil cluster santri tersebut pengukuran jarak menggunakan manhattan lebih baik dibanding dengan minkowski dimana nilai DBI manhattan sebesar -0,0869.

**Kata Kunci:** clustering, data mining, k-medoids, manhattan, minkowski

## *Implementation of Data Mining for Clustering Students Using the K-Medoids Method*

### *Abstract*

*Clustering is a process of grouping data, observations or grouping classes that have the same objects. Daarul Ilmi boarding school Islamic boarding school must also group or cluster its students to maximize the teaching process. The Islamic boarding school does not have a computer application for grouping students so that there are often mistakes in placing students according to the criteria that have been set by the Islamic boarding school. The process of grouping students is carried out based on the value of ahlak, worship, average report card scores, achievements and tahfidz scores. The grading process is still subjective, meaning that the grading is based on the grading of teachers who are effective in certain subjects, so this research aims to help determine the decision to categorize students objectively. This study uses the K-Medoids Algorithm or commonly known as partitioning around method (PAM). Based on the trial of the program carried out, this study succeeded in grouping students into murtafi and mumtaz classes. Based on the calculation of distance using manhattan, 15 students are included in the murtafi class and 39 students are included in the mumtaz class, while if using the Minkowski distance calculation, 14 students are included in the apostasy class and 40 students are included in the mumtaz class. From the evaluation using the Davies Bouldin Index (DBI) on the results of the student cluster, the distance measurement using manhattan is better than minkowski where the DBI value of manhattan is -0.0869.*

**Keywords:** clustering, data mining, k-medoids, manhattan, minkowski

## I. PENDAHULUAN

Perkembangan teknologi yang semakin pesat di era globalisasi saat ini tidak bisa dihindari lagi pengaruhnya terhadap dunia Pendidikan. Tuntutan global menuntut dunia Pendidikan untuk selalu dan senantiasa menyesuaikan perkembangan teknologi terhadap usaha dalam peningkatan mutu Pendidikan, terutama penyesuaian penggunaan teknologi dan komunikasi bagi dunia pendidikan khususnya dalam proses pembelajaran.

Pondok Pesantren Daarul Ilmi Muhammadiyah Boarding School Tarakan merupakan salah satu sekolah swasta yang terletak di Jl. Jendral Sudirman Pamusian yang melakukan pengelompokan santri pada asrama berdasarkan kriteria tertentu. Pengelompokan asrama santri ini dikhususkan bagi siswa kelas IX yang bertujuan untuk dapat lebih mudah mengetahui santri yang masuk kategori kelas murtafi atau mumtaf dan metode k-medoids dari teknik clustering dapat membantu dalam pengelompokan santri. Siswa kelas IX saat ini berjumlah 54 orang. Dengan jumlah tersebut dan beberapa atribut, maka dibutuhkan sebuah sistem terkomputerisasi untuk memudahkan proses pengelompokan santri.

Pada penelitian ini penulis menggunakan teknik *Clustering* dengan algoritma *K-Medoids* dan menggunakan perhitungan jarak manhattan distance dan minkowski distance. Algoritma *K-Medoids* atau yang biasa disebut dengan PAM (Partitioning Around Medoids) adalah metode partisi untuk mengelompokkan sekumpulan  $n$  objek menjadi sejumlah  $k$  cluster [1]. Berdasarkan penelitian ini metode yang digunakan yaitu *K-Medoids* nantinya akan menghasilkan aplikasi yang digunakan untuk mengelompokkan asrama santri.

Algoritma *K-Medoids* dipilih karena menggunakan medoid sebagai centroid sehingga tahan terhadap outlier dan lebih efektif dalam mengelompokkan data yang tidak terlalu terkait dengan cluster lainnya [2]. Penelitian menggunakan algoritma *K-Medoids* telah dilakukan sebelumnya oleh Wawan Kurniawan dengan judul analisis algoritma *K-Medoids* clustering dalam menentukan pemesanan hotel [3]. Algoritma *K-Medoids* juga telah digunakan oleh Evi Rahmah untuk melakukan clustering dalam penentuan strategi promosi pada data mahasiswa, dimana penelitian ini bertujuan untuk menggali data mahasiswa agar tepat dalam melakukan promosi bagi STIKES perintis Padang [4]

## II. METODOLOGI PENELITIAN

Data mining merupakan proses mencari pola atau informasi menarik dalam data terpilih atau disebut juga suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database dengan menggunakan metode atau teknik tertentu. Data mining adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning dalam mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dengan berbagai basis data berkapasitas besar. Salah satu metode yang berperan penting dalam mengeksplorasi data, menciptakan prediksi dan mengatasi anomali data [5].

Dalam data mining selain klasifikasi dan asosiasi terdapat pula teknik clustering. Clustering merupakan suatu proses pengelompokan data, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Salah satu algoritma yang populer dalam pengaplikasian metode partisi ini yaitu algoritma *K-Means* dan algoritma *K-Medoids*.

Adapun proses clustering menggunakan *k-medoids* adalah sebagai berikut:

1. Tentukan  $k$  (jumlah cluster) yang diinginkan
2. Pilih secara acak medoid awal sebanyak  $k$  dari  $n$  data
3. Hitung jarak masing-masing objek ke medoids sementara, kemudian tandai jarak terdekat objek ke medoids dan hitung totalnya.
4. Lakukan iterasi *k-medoids*.
5. Hitung total simpangan ( $S$ )  
Jika  $a$  adalah jarak terdekat antara objek ke medoid awal, dan  $b$  adalah jumlah jarak terdekat antara objek ke medoids baru, maka total simpangan adalah  $S = b - a$ .
6. Ulangi Langkah 3 sampai 5 dan hentikan jika sudah tidak terjadi perubahan anggota medoids.

Pada penelitian ini digunakan dua metode pengukuran jarak yaitu menggunakan manhattan distance dan minkowski distance.

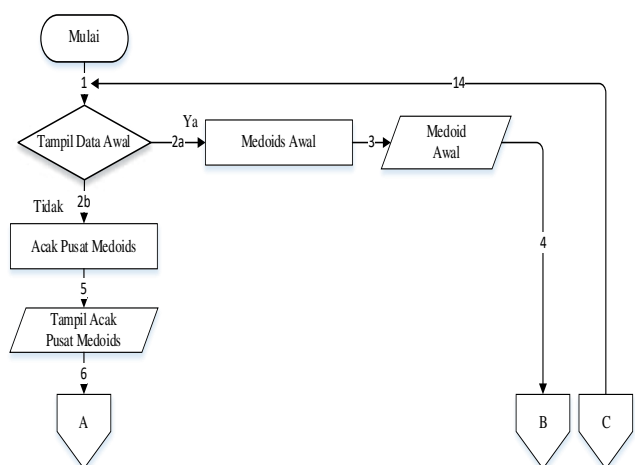
Manhattan distance atau jarak manhattan distance sering juga disebut dengan taxicab distance atau city block distance digunakan untuk menghitung jarak dengan tujuan untuk mendapatkan jarak dari satu titik data ke titik data yang lain [6] [7].

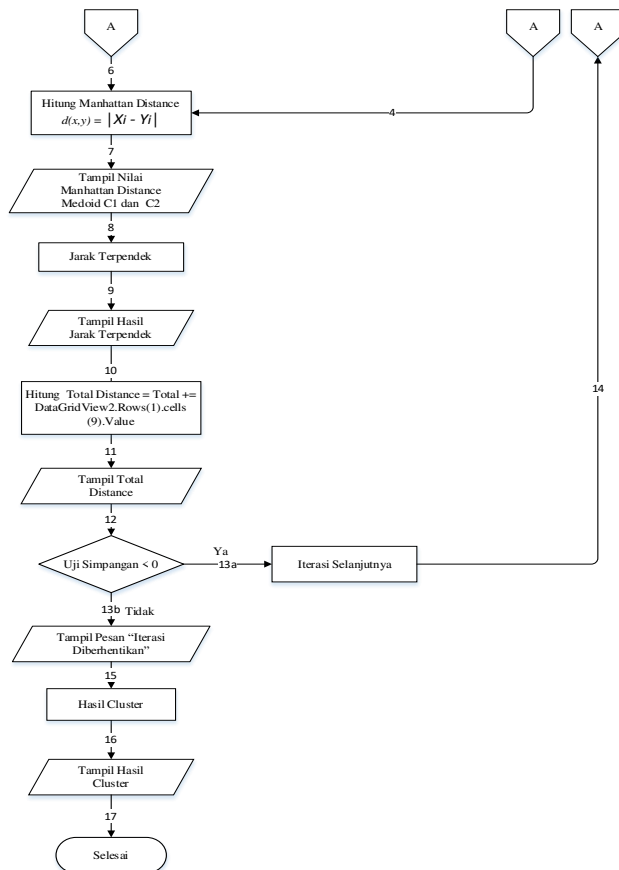
Jarak manhattan dihitung sebagai jumlah dari perbedaan mutlak antara dua vektor. Berikut ini adalah rumus manhattan distance:

$$d(x,y) = \sum_{i=1}^n |X_i - Y_i| \quad (1)$$

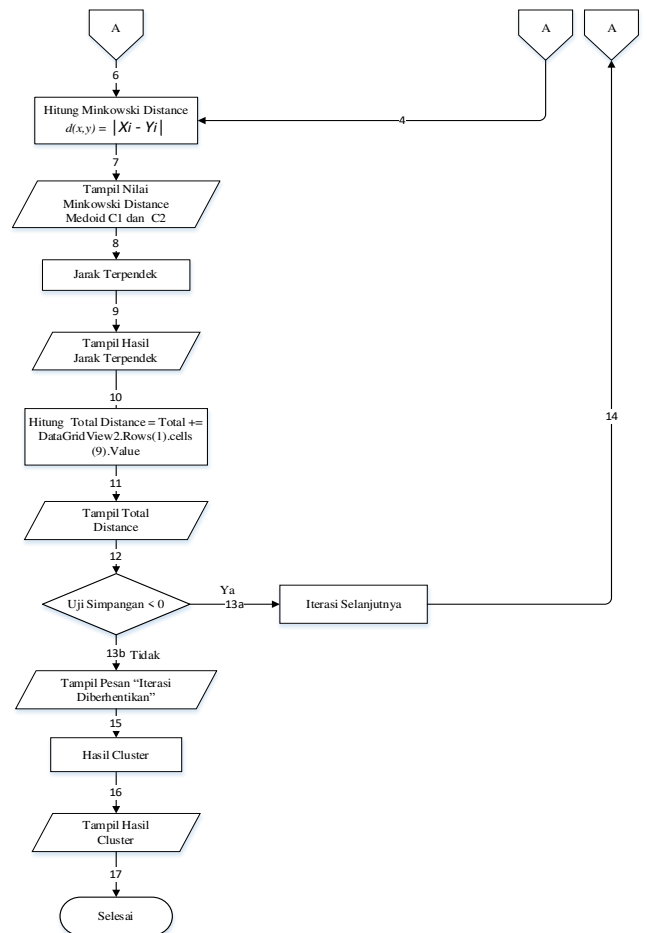
Dimana  $d$  merupakan jarak  $x$  dan  $y$  ( $x$  merupakan pusat cluster,  $x$  adalah nilai pada atribut,  $i$  menyatakan setiap data,  $n$  jumlah data,  $x_i$  merupakan nilai pada pusat cluster dan  $y_i$  menyatakan nilai pada data ke  $i$ ).

Adapun flowchart pada saat clustering santri menggunakan Manhattan distance dapat dilihat pada gambar 1 sebagai berikut :





Gambar 1. Flowchart Clustering Manhattan Distance

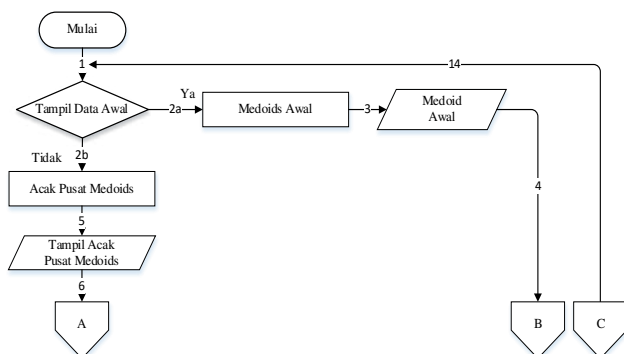


Gambar 2. Flowchart Clustering Minkowski Distance

Selanjutnya Minkowski distance adalah matrik ukur yang digunakan untuk menghitung jarak antara dua vektor bernilai bilangan riil, sekaligus dianggap sebagai generalisasi dari Euclidean distance dan Manhattan distance. Dalam pengukuran jarak objek menggunakan minkowski distance biasanya digunakan nilai  $p$  yang mana " $p$ " adalah parameter order. Berikut ini adalah rumus minkowski distance [8] [9].

$$d(x,y) = (\sum_{i=1}^n |Xi - Yi|^p)^{1/p} \quad (2)$$

Dimana  $d$  merupakan jarak  $x$  dan  $y$  ( $x$  merupakan pusat cluster,  $x$  adalah nilai pada atribut,  $i$  menyatakan setiap data,  $n$  jumlah data,  $xi$  merupakan nilai pada pusat cluster dan  $yi$  menyatakan nilai pada data ke -  $i$  sedangkan  $p$  adalah parameter order. Sedangkan untuk pengukuran jarak menggunakan minkowski distance flowchart dapat dilihat pada gambar 2 sebagai berikut :



### III. HASIL DAN PEMBAHASAN

Pada penelitian ini akan dibangun sistem sebagai implementasi metode clustering dalam pengelompokkan santri berdasarkan nilai. Hasil yang akan keluar adalah cluster yang dibagi menjadi 2 yaitu kelas mumtaz dan murtafi.

Penerapan algoritma K-Medoids pada proses pengelompokkan santri berdasarkan nilai santri. Berikut data nilai santri yang datanya diambil dari penilaian santri kelas IX sebanyak 54 data yang akan dikelompokkan menjadi 2 cluster. Atribut yang digunakan yaitu nilai akhlak, ibadah,

rata-rata nilai raport, prestasi dan nilai tahfidz. Adapun contoh data santri dapat dilihat pada tabel 1.

Tabel 1. Data Santri

| Kode Santri | Nama Santri    | Akhlak | Ibadah | Rata-Rata | Prestasi | Tahfidz |
|-------------|----------------|--------|--------|-----------|----------|---------|
| S001        | Adinda Salsa   | 80     | 75     | 82        | 0        | 65      |
| S002        | Al Huzni       | 90     | 90     | 77        | 0        | 39      |
| S003        | Alifia H.      | 90     | 85     | 88        | 0        | 67      |
| S004        | Alma Xena Z    | 86     | 80     | 85        | 20       | 65      |
| S005        | Amelia F.      | 90     | 90     | 82        | 0        | 75      |
| S006        | Annisa Sri H   | 86     | 80     | 86        | 20       | 93      |
| S007        | Aulia Citra C  | 80     | 80     | 83        | 30       | 80      |
| S008        | Aulia Fitriani | 85     | 85     | 88        | 0        | 91      |
| ....        | ....           | ....   | ....   | ....      | ....     | ....    |
| S054        | Rofi'ul H.M.   | 87     | 90     | 85        | 20       | 99      |

Dari tabel 1, Proses cluster untuk penilaian santri berdasarkan beberapa atribut dapat dilakukan dengan langkah-langkah sebagai berikut:

#### A. Perhitungan Jarak Manhattan Distance

1. Menginisialisasi pusat cluster sebanyak k (jumlah cluster) menentukan nilai centroid awal yang dipilih secara random. Diasumsikan bahwa siswa bernama Kenia Rayah dan Nur Rahma Maimunah adalah medoids. Jadi pusat medoids 1 yaitu : (90, 90, 88, 20, 100) dan pusat medoids 2 yaitu : (80, 80, 77, 0, 66). Data awal dapat dilihat pada tabel 2.

Tabel 2. Medoids Awal

| Nama Santri            | Akhlak | Ibadah | Rata-Rata Nilai Raport | Prestasi | Tahfidz |
|------------------------|--------|--------|------------------------|----------|---------|
| Kenia Rayah Alfakhirah | 90     | 90     | 88                     | 20       | 100     |
| Nur Rahma Maimunah     | 80     | 80     | 77                     | 0        | 66      |

2. Menghitung jarak menggunakan Manhattan Distance dari data pada tabel 1, setelah menentukan pusat medoids 1 dan pusat medoids 2 seperti yang terlihat pada tabel 2. Berikut adalah salah satu contoh perhitungan untuk mendapatkan hasil dari persamaan 1.

Medoids 1 :

D (Kenia Rayah Alfakhirah)

$$= |80-90| + |75-90| + |82-88| + |0-20| + |65-100|$$

$$= 10 + 15 + 6 + 20 + 35$$

$$= 86$$

Medoids 2 :

D (Nur Rahma Maimunah)

$$= |80-80| + |75-80| + |82-77| + |0-0| + |65-6|$$

$$= 0 + 5 + 5 + 0 + 1$$

$$= 11$$

Hasil dari perhitungan pengukuran jarak menggunakan Manhattan Distance medoids 1 adalah (86) dan medoids 2 (11) seperti yang terlihat pada tabel 3.

Tabel 3. Perhitungan Manhattan Distance Iterasi Pertama

| Kode Santri | Akhlak | Ibadah | .... | Cost1             | Cost 2 | Jarak terdekat | Cluster |
|-------------|--------|--------|------|-------------------|--------|----------------|---------|
| S001        | 80     | 75     | ...  | 86                | 11     | 11             | C2      |
| S002        | 90     | 90     | ...  | 92                | 47     | 47             | C2      |
| S003        | 90     | 85     | ...  | 58                | 27     | 27             | C2      |
| S004        | 86     | 80     | ...  | 52                | 35     | 35             | C2      |
| S005        | 90     | 90     | ...  | 51                | 34     | 34             | C2      |
| S006        | 86     | 80     | ...  | 23                | 62     | 23             | C1      |
| S007        | 80     | 80     | ...  | 55                | 50     | 50             | C2      |
| S008        | 85     | 85     | ...  | 39                | 46     | 39             | C1      |
| ....        | ....   | ....   | .... | ....              | ....   | ....           | ....    |
| S054        | 87     | 90     | ...  | 7                 | 78     | 7              | C1      |
|             |        |        |      | <b>Total Cost</b> |        | <b>1495</b>    |         |

3. Memilih objek pusat medoids baru yang ditentukan secara random. Diasumsikan pusat medoids baru 1 dan pusat medoids baru 2 adalah santri bernama Aziza Tul'muminah dan Messi Adesta Raysa F. Jadi pusat medoids baru 1 = (78, 78, 78, 0, 47) dan pusat medoids baru 2 = (90, 85, 88, 20, 93). Data medoids baru dapat dilihat pada tabel 4.

Tabel 4. Medoids Baru

| Nama Santri        | Akhlak | Ibadah | Rata-Rata Nilai Raport | Prestasi | Tahfidz |
|--------------------|--------|--------|------------------------|----------|---------|
| Aziza Tul Mu'minah | 78     | 78     | 78                     | 0        | 47      |
| Messi Adesta R     | 90     | 85     | 88                     | 20       | 93      |

4. Perhitungan dilakukan kembali menggunakan persamaan 2.1 dan menggunakan medoids baru yang telah ditentukan pada tabel 4, adapun hasil perhitungan pada iterasi ke-2 dapat dilihat pada tabel 5.

Tabel 5. Iterasi Kedua Perhitungan Manhattan

| Kode Santri     | Akhlak | Ibadah | .... | Cost1 | Cost 2 | Jarak terdekat | Cluster |
|-----------------|--------|--------|------|-------|--------|----------------|---------|
| S001            | 80     | 75     | ...  | 27    | 74     | 27             | C1      |
| S002            | 90     | 90     | ...  | 33    | 90     | 33             | C1      |
| S003            | 90     | 85     | ...  | 49    | 46     | 46             | C2      |
| S004            | 86     | 80     | ...  | 55    | 40     | 40             | C2      |
| S005            | 90     | 90     | ...  | 56    | 49     | 49             | C2      |
| S006            | 86     | 80     | ...  | 84    | 11     | 11             | C2      |
| S007            | 80     | 80     | ...  | 72    | 43     | 43             | C2      |
| S008            | 85     | 85     | ...  | 68    | 27     | 27             | C2      |
| ....            | ....   | ....   | .... | ....  | ....   | ....           | ....    |
| S054            | 87     | 90     | ...  | 100   | 17     | 17             | C2      |
| Total Cost Baru |        |        |      |       |        |                | 1739    |

Hasil dari perhitungan manhattan distance medoids 1 adalah (27) dan medoids 2 (74).

5. Melakukan perhitungan total simpangan (S) sebagai berikut:

$$\begin{aligned}
 S &= \text{Total Cost Baru} - \text{Total Cost Lama} \\
 &= 1739 - 1495 \\
 &= 244
 \end{aligned}$$

Karena nilai  $S > 0$  maka proses pengklasteran dihentikan. Sehingga diperoleh anggota tiap cluster.

6. Karena total cost yang baru lebih besar dari pada total cost lama, maka perhitungan dihentikan, sehingga didapatkan cluster beserta anggota masing-masing cluster. Hasil cluster dapat dilihat pada tabel 6.

Tabel 6. Hasil Clustering

| No. | Hasil Cluster 1 dan 2     |                      |
|-----|---------------------------|----------------------|
|     | Hasil Cluster 1           | Hasil Cluster 2      |
| 1   | Annisa Sri Hidayat        | Adinda Salsa R       |
| 2   | Aulia Fitriani            | Al Huzni             |
| 3   | Fairuz Athiyyah Salsabila | Alifia Hilmiya       |
| 4   | Fyersha Ayunita           | Alma Xena Zulhijjah  |
| 5   | Ghaida Aliyyah Ardiana    | Amelia Fitriana      |
| 6   | Kenia Rayah Alfakhirah    | Aulia Citra Chantika |
| 7   | Kirana Sakhi Fatimah S.   | Haya Tsanawi         |
| 8   | Messi Adesta Raysa F.     | Nadien Nur Fitriah   |
| 9   | Nailatul Hidayah          | Naya Azisah          |
| 10  | Nur Aina Ramadhani        | Nur Aulia            |
| 11  | Nur Azmi Afifah           | Nur Fadhliah M.R.    |
| 12  | Nurul Hidayah             | Nur'annisa           |
| 13  | Shafna Fithriyana         | Rayza Syahra Tunnisa |
| 14  | Siti Shofiyyah            | Shofiyah Al Balqis   |
| 15  | Tsurayya                  | Siti Adawiyah QAP    |
| 16  | Utami Anggun Pratiwi      | Siti Fatimah Sarwan  |
| 17  | Zaskiyah Salsabila A.     | Zainab Nur Sa'idah   |
| 18  | Adzara Dzakiyyah          | Agnes Syavina Candy  |
| 19  | Azka Amany Syahana        | Astri Ramadhani      |
| 20  | Indah Mutiara Suhardi     | Aulia Eka Septi R    |

|    |                           |                       |
|----|---------------------------|-----------------------|
| 21 | Muzdalifah Aulia Fadillah | Aziza Tul Mu'minah    |
| 22 | Nur Badilah Afiyah        | Davila Hijriah Ridani |
| 23 | Rofi'ul Husna Masythoh    | Diva Fiantika         |
| 24 |                           | Fairis Althafunnisa   |
| 25 |                           | Febby Aulia Fadillah  |
| 26 |                           | Hazimah Nurul H       |
| 27 |                           | Husnul Khotimah       |
| 28 |                           | Marsya Quria N        |
| 30 |                           | Nurul Aisyiyah Qirani |
| 31 |                           | Putri Handayani       |

Dari tabel 6 dapat disimpulkan bahwa kelas murtafi (cluster 1) terdiri dari 23 santri dan kelas mumtaz (cluster 2) terdiri dari 31 santri.

#### B. Perhitungan Jarak Minkowski Distance

1. Menginisialisasi pusat cluster sebanyak k (jumlah cluster) menentukan nilai centroid awal yang dipilih secara random. Diasumsikan bahwa santri bernama Kenia Rayah Alfakhri dan Nur Rahma Maimunah adalah medoids. Jadi pusat medoids 1 yaitu = (90, 90, 88, 20, 100), pusat medoids 2 yaitu = (80, 80, 77, 0, 66). Data awal dapat dilihat pada tabel 7.

Tabel 7. Medoids Awal

| Nama Santri            | Akhlak | Ibadah | Rata-Rata Nilai Raport | Prestasi | Tahfidz |
|------------------------|--------|--------|------------------------|----------|---------|
| Kenia Rayah Alfakhirah | 90     | 90     | 88                     | 20       | 100     |
| Nur Rahma Maimunah     | 80     | 80     | 77                     | 0        | 66      |

2. Menghitung jarak menggunakan metode minkowski distance dari data yang ada pada tabel 1 dengan menggunakan medoids awal pada tabel 7. Melalui persamaan 2, berikut salah satu contoh perhitungan jarak menggunakan metode minkowski.

Medoids 1 :

D (Kenia Rayah Alfakhirah)

$$\begin{aligned}
 &= (|80-90|^2 + |75-90|^2 + |82-88|^2 + |0-20|^2 + |65-100|^2)^{1/2} \\
 &= 100 + 225 + 36 + 400 + 1225 \\
 &= 44,56
 \end{aligned}$$

Medoids 2 :

D (Nur Rahma Maimunah)

$$\begin{aligned}
 &= (|80-80|^2 + |75-80|^2 + |82-77|^2 + |0-0|^2 + |65-66|^2)^{1/2} \\
 &= 0 + 25 + 25 + 0 + 1 \\
 &= 7,14
 \end{aligned}$$

Hasil dari Minkowski Distance medoids 1 adalah (44,56) dan medoids 2 (7,14), Seperti yang terlihat pada tabel 8.

Tabel 8. Perhitungan Minkowski Iterasi Pertama

| Kode Santri | Akhlak | Ibadah | .... | Cost1 | Cost 2 | Jarak terdekat | Cluster |
|-------------|--------|--------|------|-------|--------|----------------|---------|
| S001        | 80     | 75     | ...  | 44,56 | 7,14   | 7,14           | C2      |
| S002        | 90     | 90     | ...  | 65,13 | 30,48  | 30,48          | C2      |

|                   |      |      |      |       |       |       |               |
|-------------------|------|------|------|-------|-------|-------|---------------|
| S003              | 90   | 85   | ...  | 38,91 | 15,72 | 15,72 | C2            |
| S004              | 86   | 80   | ...  | 36,74 | 22,38 | 22,38 | C2            |
| S005              | 90   | 90   | ...  | 32,57 | 17,49 | 17,49 | C2            |
| S006              | 86   | 80   | ...  | 13,00 | 35,30 | 13,00 | C1            |
| S007              | 80   | 80   | ...  | 26,93 | 33,65 | 26,93 | C1            |
| S008              | 85   | 85   | ...  | 23,04 | 28,21 | 23,04 | C1            |
| ....              | .... | .... | .... | ....  | ....  | ....  | ....          |
| S054              | 87   | 90   | .... | 4,36  | 41,26 | 4,36  | C1            |
| <b>Total Cost</b> |      |      |      |       |       |       | <b>958,73</b> |

3. Memilih objek pusat medoids baru yang ditentukan secara random. Diasumsikan pusat medoids baru 1 dan pusat medoids baru 2 adalah santri bernama Aziza Tul'muminah dan Messsi Adesta Raysa F. Jadi pusat medoids baru 1 = (78, 78, 0, 47) dan pusat medoids baru 2 = (90,85,88,20,93). Data medoids baru dapat dilihat pada tabel 9.

Tabel 9. Medoids Baru

| Nama Santri           | Ahlak | Ibadah | Rata-Rata Nilai Raport | Prestasi | Tahfidz |
|-----------------------|-------|--------|------------------------|----------|---------|
| Aziza Tul Mu'minah    | 78    | 78     | 78                     | 0        | 47      |
| Messi Adesta Raysa F. | 90    | 85     | 88                     | 20       | 93      |

4. Perhitungan dilakukan ulang menggunakan persamaan 2.2 dan menggunakan medoids baru yang telah ditentukan pada tabel 9.

Medoids 1 baru :

D (Aziza Tul'muminah

$$= (|80-78|^2 + |75-78|^2 + |82-78|^2 + |0-0|^2 + |65-47|^2)^{1/2}$$

$$= 4 + 9 + 16 + 0 + 324$$

$$= 18,79$$

Medoids 2 baru :

D (Messi Adesta Rayza F. )

$$= (|80-90|^2 + |75-85|^2 + |82-88|^2 + |0-20|^2 + |65-93|^2)^{1/2}$$

$$= 100 + 100 + 36 + 400 + 784$$

$$= 37,68$$

Hasil dari Minkowski Distance medoids 1 adalah (18,79) dan medoids 2 (37,68) . Seperti yang terlihat pada tabel 10.

Tabel 10. Perhitungan Iterasi Kedua Minkowski

| Kode Santri | Akhlak | Ibadah | .... | Cost1 | Cost 2 | Jarak terdekat | Cluster |
|-------------|--------|--------|------|-------|--------|----------------|---------|
| S001        | 80     | 75     | ...  | 18,79 | 37,68  | 18,79          | C1      |
| S002        | 90     | 90     | ...  | 18,79 | 58,84  | 18,79          | C1      |
| S003        | 90     | 85     | ...  | 26,32 | 32,80  | 26,32          | C1      |
| S004        | 86     | 80     | ...  | 29,00 | 28,88  | 28,88          | C2      |
| S005        | 90     | 90     | ...  | 32,98 | 28,02  | 28,02          | C2      |

|                        |      |      |      |       |       |       |                |
|------------------------|------|------|------|-------|-------|-------|----------------|
| S006                   | 86   | 80   | ...  | 51,46 | 6,71  | 6,71  | C2             |
| S007                   | 80   | 80   | ...  | 44,97 | 20,47 | 20,47 | C2             |
| S008                   | 85   | 85   | ...  | 46,20 | 20,71 | 20,71 | C2             |
| ....                   | .... | .... | .... | ....  | ....  | ....  | ....           |
| S054                   | 87   | 90   | ...  | 58,12 | 8,89  | 19,24 | C1             |
| <b>Total Cost Baru</b> |      |      |      |       |       |       | <b>1055,65</b> |

5. Perhitungan total simpangan (S)

$$S = \text{Total Cost Baru} - \text{Total Cost Lama}$$

$$= 1055,65 - 958,73$$

$$= 96,92$$

Karena nilai  $S > 0$  maka proses pengklasteran dihentikan. Sehingga diperoleh anggota tiap cluster.

6. Karena total cost yang baru lebih besar dari pada total cost lama, maka perhitungan dihentikan. Sehingga didapatkan cluster beserta anggota cluster masing-masing. Hasil cluster dapat dilihat pada tabel 11.

Tabel 11. Hasil Clustering Minkowski

| No. | Hasil Cluster 1 dan 2     |                        |
|-----|---------------------------|------------------------|
|     | Hasil Cluster 1           | Hasil Cluster 2        |
| 1   | Annisa Sri Hidayat        | Adinda Salsa Ramadhani |
| 2   | Aulia Citra Chantika      | Al Huzni               |
| 3   | Aulia Fitriani            | Alifia Hilmiya         |
| 4   | Fairuz Athiyyah Salsabila | Alma Xena Zuhijjah     |
| 5   | Fyersha Ayunita           | Amelia Fitriana        |
| 6   | Ghaida Aliyyah Ardiana P  | Haya Tsanawi           |
| 7   | Kenia Rayah Alfakhirah    | Nadien Nur Fitriah     |
| 8   | Kirana Sakhi Fatimah S.   | Naya Azisah            |
| 9   | Messi Adesta Raysa F.     | Nur Aina Ramadhani     |
| 10  | Nailatul Hidayah          | Nur Aulia              |
| 11  | Nur Azmi Afifah           | Nur Fadhliah M.R.      |
| 12  | Nurul Hidayah             | Nur'annisa             |
| 13  | Shafna Fithriyana         | Rayza Syahra Tunnisa   |
| 14  | Siti Shofiyyah            | Shofiyah Al Balqis     |
| 15  | Tsurayya                  | Siti Adawiyah QAP      |
| 16  | Utami Anggun Pratiwi      | Siti Fatimah Sarwan    |
| 17  | Zaskiyah Salsabila A.     | Zainab Nur Sa'idah     |
| 18  | Adzara Dzakiyyah          | Agnes Syavina Candy    |
| 19  | Azka Amany Syahana        | Astri Ramadhani        |
| 20  | Diva Fiantika             | Aulia Eka Septi Rahayu |
| 21  | Febby Aulia Fadillah Sari | Aziza Tul Mu'minah     |
| 22  | Indah Mutiara Suhardi     | Davila Hijriah Ridani  |
| 23  | Marsya Quria Nurjannah    | Fairis Althafunnisa    |
| 24  | Muzdalifah Aulia Fadillah | Hazimah Nurul Hikmah   |
| 25  | Nur Badilah Afiyah        | Husnul Khotimah        |

|    |                        |                       |
|----|------------------------|-----------------------|
| 26 | Rofi'ul Husna Masythoh | Nur Rahma Maimunah    |
| 27 |                        | Nurul Aisyiyah Qirani |
| 28 |                        | Putri Handayani       |

Dari tabel 11 dapat disimpulkan terdapat 26 santri masuk pada cluster 1 murtafi dan 28 santri masuk pada cluster 2 mumtaz. Setelah dilakukann clustering menggunakan manhattan dan minkowski distance, dilakukan evaluasi menggunakan Davies Bouldin Index (DBI) agar diketahui performansi masing-masing metode pengukuran jarak tersebut [10]. Manhattan distance mendapat nilai DBI sebesar -0,0869 sedangkan minkowski mendapatkan nilai DBI sebesar -0,3061. Walaupun nilai DBI masih bernilai negatif dimana hasil cluster belum terlalau baik, maka dari kedua metode tersebut metode manhattan masih memiliki performa yang lebih baik dari minkowski distance.

#### IV. KESIMPULAN

Berdasarkan penelitin yang telah dilakukan, dengan adanya aplikasi clustering, pihak pesantren lebih mudah dalam mengelompokkan santri sesuai dengan kriteria yang telah ditetapkan sehingga dapat terhindar dari kesalahan dalam mengelompokkan santri. Performansi kedua metode pengukuran jarak tersebut menunjukkan nilai DBI Manhattan lebih baik daripada minkowski dimana nilai DBI Manhattan sebesar -0,0869 dan minkowski sebesar -0,3061, hal ini menunjukkan performa manhattan distance lebih baik dibanding minkowski distance dalam melakukan pengelompokkan santri.

#### REFERENSI

- [1] Kaur, Noor K, Singh and Dheerendra,, "K-Medoids Clustering Algorithm A Review, International," *Journal Of Computer Application and Technology*, ISSN.2349-1841, vol. 1, no. 1, April 2014.
- [2] R. N. Ibrahim, M. N. Hayati and F. D. T. Amijaya, "Penerapan Algoritma K-Medoids pada Pengelompokan Wilayah Desa atau Kelurahan di Kabupaten Kertanegara," *Jurnal EKSPONENSIAL*, Vols. 11, No.2, pp. 153-157, 2020.
- [3] W.Kurniawan, A.Rifai, W.Gata, D.Gunawan, "Analisis Algoritma K-Medoids Clustering dalam Menentukan Pemesanan Hotel," *Swabumi*, Vol. %1 dari %28, No.2, 2020.
- [4] E.Rahmah, E.Haerani, A.Nazir and S.Ramadhani, "Penerapan Algoritma K-Medoids Clustering untuk Menentukan Strategi Promosi pada Data Mahasiswa," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. Vol. 5 No. 3, Juni 2022.
- [5] I.Kamila, U.Khairunnisa and Mustakim, "Perbandingan Algoritma K-Medoids untuk Pengelompokkan Data Transaksi Bongkar Muat di Propinsi Riau," *Jurnal*

*Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, pp. 119-125, 2019.

- [6] Y. Miftahuddin and S. Umaroh, "Perhitungan Euclidean, Haversine, dan Manhattan dalam Penentuan Posisi Karyawan," 2020.
- [7] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, Vols. 19, No.2, pp. 141-155, 2022.
- [8] Trivusi, "Pengertian dan Jenis-jenis Distance Metric pada Machine Learning," [Online]. Available: <https://www.trivusi.web.id/2022/06/jenis-distance-metric.html>. [Accessed 3 November 2022].
- [9] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering Berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, Vols. 4, No.1, pp. 20-24, 2019.
- [10] Warisa and Nurahman, "Perbandingan Performa Cluster Model Algoritma K-Means dalam Mengelompokkan Penerima Bantuan Program Keluarga Harapan," *Jurnal Sistem Informasi Bisnis*, pp. 20-28, 2023.