

# Clustering Bidang Keilmuan Menggunakan Kombinasi Metode Topsis dan Algoritma *K-Means*

Hani Nafisah Amaliya<sup>1</sup>, Yuni Yamasari<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

<sup>1</sup>[hani.19070@mhs.unesa.ac.id](mailto:hani.19070@mhs.unesa.ac.id)

<sup>2</sup>[yuniyamasari@unesa.ac.id](mailto:yuniyamasari@unesa.ac.id)

**Abstrak**— Banyak orang tua dan guru saat ini semakin tanggap terhadap permasalahan belajar anak. Setiap anak memiliki keistimewaan dan perangai yang berbeda. Sehingga menyebabkan pola belajar dan pola asuh diterapkan berbeda pada setiap anak. Sebagian besar anak-anak memiliki kelemahan di suatu bidang. Namun, anak juga memiliki kelebihan di bidang yang lain. Oleh karena itu, pengelompokan bidang keilmuan dengan penerapan metode TOPSIS dan algoritma *K-Means* penting dilakukan guna menentukan pola belajar yang cocok bagi anak berdasarkan bidang keilmuan yang dikuasai. Penelitian ini melakukan pengelompokan atau *clustering* bidang keilmuan berdasarkan mata pelajaran menggunakan kombinasi metode TOPSIS dan algoritma *K-Means*. Kemudian, pengukuran kinerja atau performa dilakukan terhadap model yang menerapkan kombinasi antara metode TOPSIS dengan algoritma *K-Means*. Hasil penelitian menunjukkan bahwa metode TOPSIS sebagai metode dalam penjumlahan terbobot mampu memengaruhi kinerja *clustering* bidang keilmuan berdasarkan mata pelajaran dengan algoritma *K-Means* menjadi lebih tinggi. Sehingga kombinasi metode TOPSIS dan algoritma *K-Means* terbukti lebih optimal dalam pengelompokan bidang keilmuan berdasarkan mata pelajaran dan dapat digunakan sebagai acuan dalam menentukan metode pembelajaran yang tepat berdasarkan bidang keilmuan yang dikuasai siswa.

**Kata Kunci**— TOPSIS, *K-Means*, *Clustering*, *Machine Learning*, Python.

## I. PENDAHULUAN

Anak dengan rentang usia 3 hingga 6 tahun merupakan anak usia dini, dalam masa ini anak mengalami tumbuh kembang yang pesat. Sehingga pada usia ini disebut juga dengan *Golden Age* atau masa keemasan. Anak mulai sensitif dalam menerima stimulus dan upaya mendukung perkembangan potensi anak. Masa ini adalah kesempatan dan momentum yang penting dalam merealisasikan potensi pada diri anak.

Setiap anak dilahirkan dengan karakteristik dan keunikan masing-masing. Hal ini menyebabkan pola asuh dan pola belajar yang berbeda untuk setiap anak. Pada dasarnya anak memiliki kemampuan yang berbeda dalam menangkap dan menerima pembelajaran, termasuk pola berpikir kreatif dan produktif. Anak-anak di usia ini telah mengalami perkembangan kemampuan bahasa, sehingga anak mampu untuk mengungkapkan keinginannya.

Belakangan ini banyak guru dan orang tua yang semakin peka terhadap masalah kesulitan belajar anak. Kesulitan belajar merupakan gangguan saraf yang mempengaruhi kemampuan seseorang untuk menyimpan, memproses, atau memproduksi informasi. Gangguan tersebut menciptakan kesenjangan antara

potensi intelektual seseorang dengan prestasinya. Anak-anak yang mengalami kesulitan belajar ini, biasanya mempunyai satu bidang yang menonjol dan kelemahan di bidang yang lain. Tes adalah salah satu metode yang umum digunakan untuk mengumpulkan informasi. Lebih banyak informasi tentang profil belajar anak, akan lebih mempermudah dalam merencanakan pendidikannya.

Taman Pendidikan Anak Sholeh (TAPAS) merupakan sebuah sistem pembinaan sepanjang usia (*long live education*) dengan program yang berkesinambungan dan berdampingan dengan pendidikan formal yang sudah ada. TAPAS adalah salah satu solusi penyelenggaraan layanan pendidikan bagi anak usia dini yang banyak menekankan pada pembinaan kecerdasan mental spiritual (aqidah), di samping itu juga memberikan berbagai bentuk materi pembelajaran yang lain (umum)[1].

TAPAS Al Manshuroh Jombang membutuhkan pengelompokan anak didik berdasarkan mata pelajaran yang dikuasai. Sehingga dalam penelitian ini diterapkan metode TOPSIS dan algoritma *K-Means* untuk dapat menyelesaikan permasalahan yang terjadi di TAPAS Al Mashuroh tersebut.

Penerapan *data mining* pada penelitian ini adalah untuk *clustering* bidang keilmuan yang dikuasai oleh anak usia 3-6 tahun di TAPAS Al Manshuroh Jombang. *Data mining* merupakan serangkaian proses untuk mencari dan mengumpulkan data melalui eksplorasi, seleksi, dan pemodelan dari berbagai bagian data untuk mendapatkan pola [2]. *Clustering* adalah salah satu data mining yang bersifat *unsupervised* [3].

Algoritma yang digunakan pada penelitian ini adalah algoritma *K-Means* untuk menentukan *clustering* pada data yang digunakan. Algoritma *K-Means* merupakan sebuah algoritma yang mengelompokkan data berdasarkan jarak terdekat dari suatu *cluster*, proses perhitungan akan terus dilakukan hingga tidak ada lagi pergeseran pada *cluster*.

TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) merupakan salah satu metode yang digunakan dalam penyelesaian kasus *Multi Attributes Decision Making*. TOPSIS adalah metode penjumlahan terbobot untuk mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif [4].

Penelitian serupa pernah dilakukan oleh [5] pada tahun 2018 dengan judul Pengelompokan Bidang Keilmuan Di Teknologi Informasi Dengan Metode *K-Means* Dan Optimasi *Simple Additive Weighting* (SAW) Dalam Penentuan Kesesuaian Terhadap Keilmuan. Dalam penelitiannya, *clustering* menggunakan algoritma *K-Means* dan menggunakan

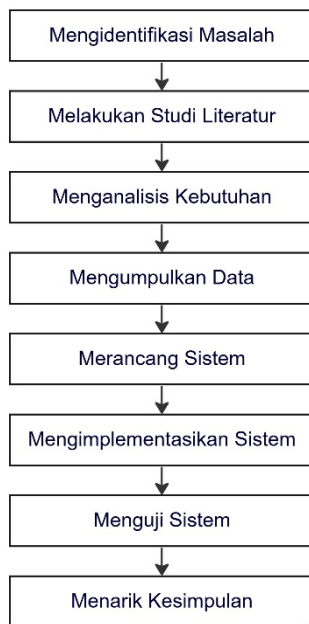
penjumlahan terbobot dengan metode SAW. Dari penelitian terdahulu terlihat bahwa penelitian tersebut bertujuan untuk melihat pengaruh dari pengukuran jarak *Euclidean Distance* dan *Manhattan Distance* pada metode TOPSIS dalam penentuan nilai solusi ideal positif dan negatif yang datanya telah dikelompokkan menggunakan algoritma *K-Means* [6].

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh [5] adalah pada proses *clustering* diterapkan kombinasi antara metode TOPSIS dan algoritma *K-Means* yaitu penjumlahan terbobot hingga perhitungan jarak menggunakan metode *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS) sedangkan untuk penentuan *centroid* dan penentuan *cluster* menggunakan algoritma *K-Means*. Hasil *clustering* untuk menentukan bidang keilmuan yang dikuasai oleh anak usia 3-6 tahun.

Dari latar belakang di atas, maka peneliti ingin melakukan penelitian terkait dengan *Clustering* Bidang Keilmuan Berdasarkan Mata Pelajaran Menggunakan Metode TOPSIS dan Algoritma *K-Means* dan mengukur performa atau kinerja kombinasi metode TOPSIS dan algoritma *K-Means*. Diharapkan pada penelitian ini dapat meng-*clustering* bidang keilmuan dengan hasil yang maksimal.

## II. METODOLOGI PENELITIAN

Metode yang diterapkan pada penelitian ini adalah metode kuantitatif. Berikut merupakan alur tahapan yang dilakukan dalam penelitian ini :



Gambar 1. Alur Penelitian

### A. Mengidentifikasi Masalah

Tahap pertama yaitu mengidentifikasi masalah, permasalahan yang menjadi fokus penyelesaian dalam

penelitian ini adalah bagaimana pengelompokan atau *clustering* bidang keilmuan berdasarkan mata pelajaran yang dikuasai anak berusia 3-6 tahun di TAPAS AI Manshuroh Jombang dengan menggunakan algoritma *K-Means* dan metode TOPSIS, sehingga anak yang memiliki karakteristik yang berbeda tidak akan mengalami kesulitan dalam belajar dengan adanya pengelompokan tersebut.

### B. Melakukan Studi Literatur

Tahap kedua yaitu melakukan studi literatur, hal ini dilakukan demi merealisasikan tujuan serta pemecahan masalah penelitian dengan mempelajari teori-teori dan menggali pengetahuan tentang *machine learning*, *k-means*, TOPSIS, dan *clustering*. Selain itu, peneliti juga menggunakan referensi lain yang meliputi karya tulis ilmiah, jurnal, *e-book*, dan buku referensi mata kuliah yang dapat menunjang penelitian ini.

### C. Menganalisis Kebutuhan

Tahap ketiga yaitu menganalisis kebutuhan untuk menentukan komponen-komponen yang dibutuhkan dalam penerapan proses pengelompokan atau *clustering* bidang keilmuan dengan menggunakan algoritma *K-Means* dan metode TOPSIS.

1. Alat Penelitian
  - a. Laptop Asus Vivobook M413I dengan prosesor AMD Ryzen 5 4000 Series dan AMD Radeon Graphics.
  - b. Sistem Operasi Windows 10 64-bit.
  - c. Jupyter Notebook 6.0.3
  - d. Python 3.8.3
  - e. Pandas 1.0.5
  - f. NumPy 1.18.5
  - g. Scikit-learn 1.1.2
2. Bahan Penelitian
  - a. Artikel tentang metode TOPSIS.
  - b. Artikel tentang algoritma *K-Means*.

### D. Mengumpulkan Data

Tahap keempat yaitu mengumpulkan data, teknik pengumpulan data yang digunakan dalam penelitian ini adalah dengan mendatangi langsung TAPAS AI Manshuroh Jombang.

#### 1. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data kuantitatif yang berupa data nilai mata pelajaran siswa TAPAS AI Manshuroh Jombang. Proses pengambilan data dilakukan dengan metode observasi dan wawancara terstruktur.

#### 2. Populasi dan Sampel

##### a. Populasi

Populasi yang digunakan oleh penulis yaitu data siswa TAPAS AI Manshuroh Jombang yang sudah ditentukan.

##### b. Sampel

Sampel dalam penelitian ini adalah data siswa TAPAS Al Manshuroh Jombang yang berusia 3-6 tahun tahun ajaran 2022/2023.

### 3. Tempat dan Waktu Penelitian

#### a. Tempat

Penelitian ini dilakukan di TAPAS Al Manshuroh Jombang yang berlokasi di Dsn. Buduran, Ds. Jogoloyo, Kec. Sumobito, Kab. Jombang.

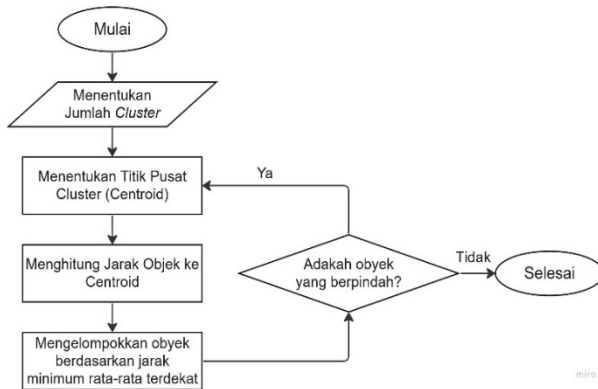
#### b. Waktu

Penelitian ini dilakukan dengan rentang waktu 26 Desember hingga 31 Desember tahun 2022.

### E. Merancang Sistem

Tahap kelima yaitu merancang sistem, sistem yang dirancang dan digunakan dalam penelitian menggunakan kombinasi metode TOPSIS dan algoritma *K-Means* untuk pengelompokan bidang keilmuan.

#### 1. K-Means



Gambar 2. Flowchart K-Means

Algoritma *K-Means* bekerja sesuai dengan banyak *K* yang telah didefinisikan dalam pengelompokan data. Algoritma ini sering disebut juga dengan *flat clustering*, dimana kedudukan satu kelompok sejajar dengan kedudukan kelompok yang lainnya [7]. Metode ini menggunakan data fisik yang jelas dan tidak abstrak, pengguna dapat menentukan *cluster* yang akan dibuat, sehingga metode ini juga bersifat fleksibel [8]. Hal ini sesuai dengan data yang akan digunakan dalam penelitian ini, yaitu dengan memanfaatkan nilai rapor siswa sebagai acuannya.

Langkah - langkah dalam menggunakan metode *K-Means* adalah sebagai berikut [9] :

- Menentukan *k* sebagai jumlah *cluster* yang akan dibentuk.
- Memilih titik pusat *cluster* (*centroid*) awal.

$$V = \frac{\sum_{i=1}^n X_i}{n}; i = 1, 2, 3, \dots, n \quad (1)$$

**Keterangan :**

*V* = *Centroid*

*X<sub>i</sub>* = Objek ke-*i*

*n* = Banyaknya objek

- Menghitung jarak setiap objek dari masing-masing *cluster* ke masing-masing *centroid*. *Euclidean Distance* digunakan dalam perhitungan jarak antara *centroid* dan objek dengan persamaan seperti berikut :

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}; i = 1, 2, \dots, n \quad (2)$$

**Keterangan :**

*X<sub>i</sub>* = Obyek X ke-*i*

*Y<sub>i</sub>* = Obyek Y ke-*i*

*n* = Banyaknya obyek

- Mengelompokkan setiap data berdasarkan jarak paling dekat antara data dengan titik pusat *cluster* (*centroid*) nya.
- Menghitung rasio untuk mengetahui iterasi selanjutnya.
- Menentukan posisi titik pusat *cluster* baru dengan cara menghitung nilai rata-rata dari setiap data yang ada pada *centroid* yang sama.

$$C_k = \left(\frac{1}{nk}\right) \sum d_i \quad (3)$$

**Keterangan :**

*C<sub>k</sub>* = Pusat *cluster* baru

*nk* = Jumlah

*d<sub>i</sub>* = Jumlah data dalam *cluster*

- Mengulang langkah ke-2 jika nilai rasio iterasi selanjutnya lebih sedikit atau sama dengan nilai rasio iterasi sebelumnya atau jika posisi *centroid* tidak sama, maka iterasi dinyatakan berhenti.

Pengecekan konvergensi dilakukan dengan cara membandingkan matriks kelompok pada iterasi sebelumnya dengan elemen matriks kelompok selama pengulangan. Apabila hasilnya sama, maka algoritma *K-Means* selesai, namun apabila berbeda, artinya belum konvergen, sehingga dilakukan pengulangan hingga konvergen [10].

#### 2. TOPSIS

Langkah - langkah dalam menggunakan metode perhitungan TOPSIS adalah sebagai berikut [4], [11] :

- Membuat normalisasi matriks keputusan

$$DM = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ \begin{matrix} L1 \\ L2 \\ \dots \\ Lm \end{matrix} & \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \end{matrix} \quad (4)$$

**Keterangan :**

*DM* = *Decision matrix* atau matriks keputusan

*C<sub>1</sub> .. C<sub>n</sub>* = Nama kolom

*L1 .. L<sub>m</sub>* = Urutan data

- X11 .. Xmn = Data sesuai urutan baris dan kolom  
b. Menghitung matriks keputusan ternormalisasi

$$NDM = R_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad (5)$$

**Keterangan :**

NDM atau  $R_{ij}$  = Matriks ternormalisasi

$X_{ij}$  = Data

m = Urutan terakhir

i = Urutan pertama

- c. Menentukan matriks ternormalisasi dengan bobot

$$V = V_{ij} = W_j \times R_{ij} \quad (6)$$

**Keterangan :**

V atau  $V_{ij}$  = Matriks ternormalisasi terbobot

$W_j$  = Bobot

$R_{ij}$  = Matriks ternormalisasi

- d. Menentukan matriks dari solusi ideal positif dan solusi ideal negatif menggunakan rumus :

$$PIS = A^+ = \{V_1^+ V_2^+ \dots V_n^+\}, \text{ dimana} \\ V_j^+ = \{(\max_i(V_{ij}) \text{ jika } j \in J), (\max(V_{ij}) \text{ jika } j \in J)\} \quad (7)$$

**Keterangan :**

PIS atau  $A^+$  = Solusi ideal positif

$V_1^+ \dots V_n^+$  atau  $V_j^+$  = Nilai maksimum setiap kolom

$$NIS = A^- = \{V_1^- V_2^- \dots V_n^-\}, \text{ dimana} \\ V_j^- = \{(\min_i(V_{ij}) \text{ jika } j \in J), (\min(V_{ij}) \text{ jika } j \in J)\} \quad (8)$$

**Keterangan :**

NIS atau  $A^-$  = Solusi ideal negatif

$V_1^- \dots V_n^-$  atau  $V_j^-$  = Nilai minimum setiap kolom

- e. Menghitung jarak nilai masing-masing alternatif dari solusi ideal positif dan solusi ideal negatif.

$$S^+ = \sqrt{\sum_{j=1}^n (V^+ - V_{ij})^2} \quad i = 1, \dots, m \quad (9)$$

**Keterangan :**

$S^+$  = Jarak masing-masing alternatif dari solusi ideal positif

$V^+$  atau  $V_{ij}$  = Solusi ideal positif

j = Matriks ternormalisasi terbobot

i = Urutan data

m atau n = Urutan data terakhir

$$S^- = \sqrt{\sum_{j=1}^n (V^- - V_{ij})^2} \quad i = 1, \dots, m \quad (10)$$

**Keterangan :**

$S^-$  = Jarak masing-masing alternatif dari solusi ideal negatif

$V^-$  atau  $V_{ij}$  = Solusi ideal negatif

j = Matriks ternormalisasi terbobot

i = Urutan data

m atau n = Urutan data terakhir

- f. Menghitung preferensi dari solusi ideal

$$C_i = \frac{S_i^-}{(S_i^+ + S_i^-)}, \quad 0 \leq C_i \leq 1 \quad (11)$$

**Keterangan :**

$C_i$  = Preferensi solusi ideal

$S_i^-$  = Jarak masing-masing alternatif dari solusi ideal negatif

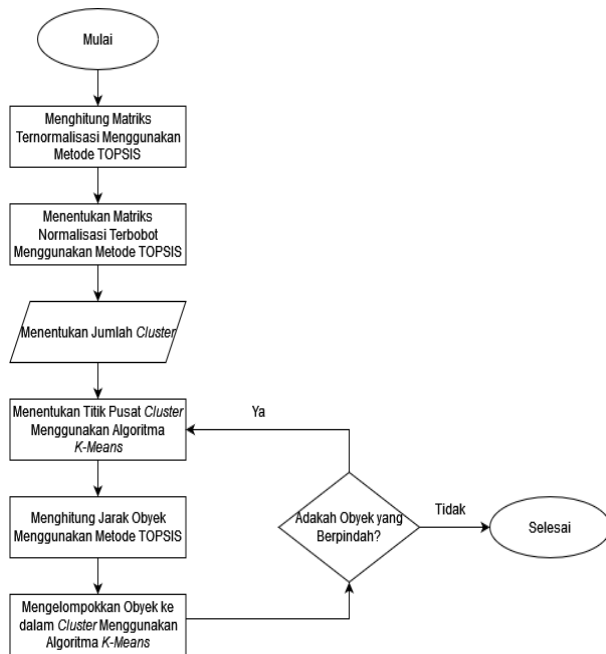
$S_i^+$  = Jarak masing-masing alternatif dari solusi ideal positif

3. Kombinasi Metode TOPSIS dan Algoritma *K-Means*

Dalam penelitian ini diterapkan kombinasi antara metode TOPSIS dengan algoritma *K-Means*, dengan langkah-langkah sebagai berikut :

- Menggunakan metode TOPSIS dalam menghitung matriks keputusan ternormalisasi sesuai dengan rumus (5)
- Menentukan matriks ternormalisasi dengan bobot dengan menggunakan metode TOPSIS sesuai dengan rumus (6)
- Menentukan jumlah *cluster* yang akan dibentuk.
- Menentukan *centroid* atau titik pusat *cluster* dengan menggunakan algoritma *K-Means*.
- Menghitung jarak masing-masing alternatif menggunakan metode TOPSIS sesuai dengan rumus (9) karena alternatif yang dipilih harus mempunyai jarak terdekat dari solusi ideal positif.
- Menggunakan algoritma *K-Means* dalam memprediksi masing-masing alternatif termasuk ke dalam *cluster* yang sesuai, berdasarkan jarak terdekat dengan *centroid*.

Berdasarkan langkah-langkah tersebut, maka alur proses dalam bentuk *flowchart* seperti gambar berikut ini :



Gambar 3. Flowchart Kombinasi TOPSIS dan K-Means

#### 4. Menghitung Kinerja

Kinerja dihitung dengan menggunakan *error*. Pada dasarnya, algoritma *K-Means* meminimalisir kesalahan atau *error* pada jumlah kuadrat *error*. Perhitungan *error* dapat diperoleh melalui persamaan seperti berikut :

$$Error = \sqrt{\sum(a - b)^2} \quad (13)$$

#### Keterangan :

a = Centroid

b = Data

#### F. Mengimplementasikan Sistem

Tahap keenam yaitu mengimplementasikan sistem dengan membuat model dari sistem yang telah dirancang pada tahapan merancang sistem sebelumnya yang dapat menyelesaikan masalah yang telah dituangkan pada bagian latar belakang. Setelah sistem diimplementasikan dan bisa dijalankan, proses selanjutnya adalah dilakukan pengujian sistem dan analisa hasil.

#### G. Menguji Sistem

Tahap ketujuh yaitu menguji sistem yang sudah diimplementasikan sebelumnya, sistem akan menguji dengan dataset yang telah ditentukan.

#### H. Menarik Kesimpulan

Tahap terakhir dari penelitian ini yaitu menarik kesimpulan dari hasil analisa data pengujian yang telah dilakukan dan

menjadi jawaban atas rumusan masalah yang telah didefinisikan sebelumnya. Kemudian penulisan saran yang bersinggungan dengan hasil yang dicapai dalam penelitian ini, sehingga dapat menjadi pertimbangan dalam penelitian selanjutnya.

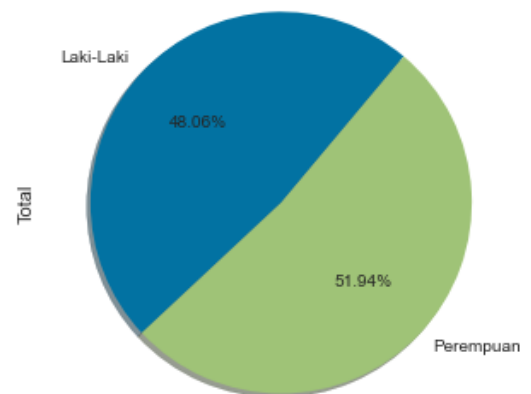
### III. HASIL DAN PEMBAHASAN

#### A. Data Pengujian

Data yang diolah oleh peneliti berupa data nilai siswa TAPAS AL Manshuroh Jombang.

##### 1. Data Siswa

Siswa TAPAS AL Manshuroh Jombang berjumlah 103 siswa, dengan jumlah siswa perempuan 53 anak dan siswa laki-laki 50 anak.



Gambar 4. Diagram Jumlah Siswa TAPAS AL Manshuroh

Terdapat 4 kelas di TAPAS AL Manshuroh Jombang, yaitu : TK A yang berjumlah 33 siswa, TK B yang berjumlah 32 siswa, PRA TK Kelompok A yang berjumlah 15 siswa, dan PRA TK Kelompok B yang berjumlah 23 siswa ditunjukkan oleh Gambar 5.

##### 2. Data Nilai

Persebaran data nilai siswa TAPAS AL Manshuroh Jombang dalam bentuk *scatter plot* dapat dilihat pada Gambar 6.

#### B. Preprocessing

Pada tahapan ini dilakukan proses mengubah data yang masih mentah menjadi data yang siap diolah dan mempermudah dalam analisis data.

##### 1. Import Library

*Library* merupakan sebuah paket atau modul yang diinstal. *Import library* bertujuan untuk dapat menjalankan fungsi-fungsi yang dibutuhkan. Beberapa *library* yang dibutuhkan dalam penelitian ini berfungsi untuk proses komputasi numerik, mengacak dan memilih

acak elemen, menampilkan berbagai jenis visualisasi, pemrosesan data, dan pemodelan *Machine Learning*.

2. *Null Value* Dataset

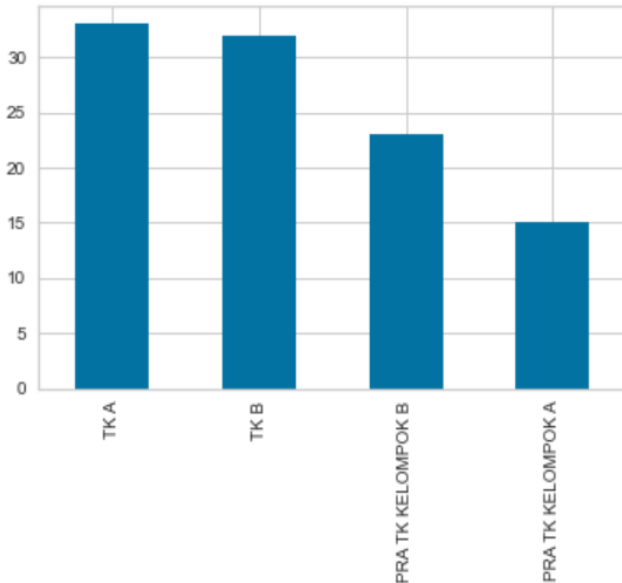
Dalam dataset berpeluang terdapat data yang masih kosong. Oleh karena itu dilakukan pengecekan *null value* dalam dataset.

3. Menghapus Kolom

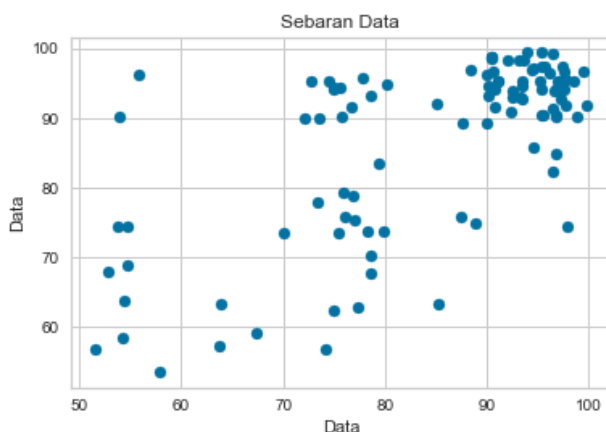
Tahapan selanjutnya yaitu menghapus beberapa kolom yang terdapat pada dataset karena tidak diperlukan dalam pemrosesan data yang digunakan untuk *clustering* bidang keilmuan.

4. *Scaling* Data

Pada penelitian ini metode data *scaler* yang digunakan



yaitu *Standard Scaler*. *Scaling* data dilakukan untuk normalisasi data.



Gambar 6. Persebaran Data Nilai

C. *TOPSIS*

1. Normalisasi

Berdasarkan rumus (5) metode TOPSIS, data ternormalisasi dengan cara setiap data dibagi dengan akar kuadrat dari jumlah masing-masing data yang telah dikuadrat sebelumnya.

2. Normalisasi Terbobot

Langkah selanjutnya setelah normalisasi data yaitu menghitung normalisasi terbobot dengan cara data ternormalisasi dikali dengan bobot pada setiap alternatif yang telah ditentukan sebelumnya, seperti pada rumus (6).

3. Solusi Ideal Positif dan Solusi Ideal Negatif

Langkah berikutnya yaitu menentukan nilai positif ideal yang didapatkan dari nilai tertinggi masing-masing data, selanjutnya menentukan nilai negatif ideal dengan mencari nilai terendah dari masing-masing data.

4. Jarak Alternatif Ideal Positif dan Ideal Negatif

Berdasarkan rumus (9) metode TOPSIS, jarak alternatif ideal positif dan ideal negatif didapatkan dengan cara mengakar kuadrat jumlah data ternormalisasi terbobot dikurangi dengan nilai positif ideal, kemudian jumlah data ternormalisasi terbobot dikurangi dengan nilai negatif ideal.

D. *K-Means*

1. Menentukan Jumlah *Cluster*

Untuk menentukan jumlah kluster, penelitian ini menggunakan metode Elbow dengan *Standard Scaler* pada rentang  $k = 1$  hingga 10. Hasil eksekusi metode memperlihatkan bahwa jumlah *cluster* yang optimal untuk digunakan yaitu  $k=3$ .

2. Menentukan *Centroid*

Langkah selanjutnya yaitu menentukan centroid atau titik pusat *cluster* awal. Penentuan *centroid* awal didapatkan dari memilih data secara acak yang ingin dijadikan sebagai titik pusat *cluster* awal.

*Centroid* awal yang didapatkan secara acak dari data yang diproses yaitu data ke-30, data ke-67, dan data ke-62. Nilai masing-masing *centroid* dapat dilihat pada Tabel I.

Setelah menentukan *centroid* awal secara acak, kemudian hitung *error* pada setiap *centroid*, nilai *error* tersebut dapat dilihat pada Tabel II.

Berdasarkan Tabel II tersebut, nilai *error* masing-masing *centroid* awal cenderung kecil, sehingga cukup bagus untuk digunakan pada tahapan selanjutnya.

3. Mengelompokkan Data

Langkah berikutnya yaitu membuat fungsi kmeans yang didalamnya berisi cara perhitungan algoritma kmeans yang mana perhitungan jaraknya dihitung dengan

menggunakan metode TOPSIS dan *centroid* awalnya berupa data acak yang telah ditentukan sebelumnya.

TABEL I  
CENTROID AWAL

Data ke-	Nilai Centroid
30	[0.32651938, 0.32804003, 0.3291382, 0.36714663, 0.30699504, 0.42269275, 0.28701366, 0.32302723, 0.47709098, 0.41918866, 0.29584768, 0.312637, 0.47604323, 0.33556009, 0.32762875, 0.52309141, 0.28623107, 0.28271885, 0.33019018, 0.31730128, 0.59267829, 0.31742869, 0.25857508, 0.312011]
67	[0.32886745, 0.32150081, 0.31905253, 0.31806204, 0.30686193, 0.45951121, 0.27284863, 0.29455002, 0.59350343, 0.44408062, 0.3464886, 0.43881974, 0.51930218, 0.33573938, 0.26162502, 0.55870268, 0.31709879, 0.3163958, 0.32601672, 0.31584013, 0.45865055, 0.29277278, 0.33212798, 0.2952967]
62	[0.31104977, 0.31883793, 0.32561837, 0.39824873, 0.31611338, 0.51120098, 0.33442524, 0.30793216, 0.51278997, 0.40314054, 0.35797641, 0.43018851, 0.48078483, 0.34481148, 0.33873481, 0.47264732, 0.31318017, 0.32024939, 0.33234646, 0.3144454, 0.49220426, 0.31558917, 0.18537294, 0.31178203]

TABEL II  
NILAI ERROR CENTROID AWAL

Centroid ke-	Nilai Error
0	0.003968743448893597
1	0.0024983504175295234
2	0.0026868767309748574

#### 4. Menentukan Centroid Baru

Posisi *centroid* dilakukan perbaruan secara berulang hingga tidak ada lagi perpindahan *centroid*. Berdasarkan proses *looping* yang telah dilakukan hingga tidak ada perpindahan data maupun perpindahan *centroid*, hasil proses *looping* dapat dilihat dalam Tabel III.

TABEL III  
NILAI CENTROID SETIAP LOOPING

Perulangan ke-	Nilai Centroid
1	[0.3071133, 0.25226589, 0.30206257, 0.31436423, 0.28113757, 0.47929964, 0.25225434, 0.26253992, 0.55880477, 0.34466181, 0.18969442, 0.3594222, 0.44415113, 0.19262971, 0.2754373, 0.42858698, 0.30125245, 0.27548082, 0.26557112, 0.28336283, 0.4326433, 0.27150544, 0.30683869, 0.27658717]

Perulangan ke-	Nilai Centroid
	[0.25131219, 0.32129856, 0.19216933, 0.31720091, 0.32157107, 0.54664025, 0.3279272, 0.33590978, 0.44620605, 0.29438104, 0.18958536, 0.29022521, 0.3654926, 0.27044178, 0.27508584, 0.45368385, 0.25147911, 0.1934498, 0.26212801, 0.32019036, 0.61419508, 0.30771477, 0.31469556, 0.30285248]
	[0.3443716, 0.30950099, 0.33055968, 0.43684775, 0.31278552, 0.59113422, 0.29354607, 0.30472045, 0.51422793, 0.45978095, 0.34241696, 0.40576604, 0.55554616, 0.26635396, 0.26120327, 0.49743126, 0.2675317, 0.33003413, 0.33585913, 0.31673675, 0.59716875, 0.29280505, 0.30645286, 0.31881446]
2	[0.26673681, 0.25824639, 0.26149834, 0.34896613, 0.27484086, 0.43948474, 0.27193285, 0.26443018, 0.44094299, 0.3298784, 0.24551962, 0.34557627, 0.4542189, 0.25200094, 0.27359268, 0.45417505, 0.28493633, 0.28450297, 0.26197625, 0.26652744, 0.4228724, 0.28073625, 0.27607234, 0.27843374]
	[0.21137204, 0.28756873, 0.23612775, 0.35203459, 0.30311809, 0.50992172, 0.30422714, 0.32656606, 0.45438576, 0.31186172, 0.22271584, 0.31819499, 0.42329027, 0.25104853, 0.24770144, 0.46628444, 0.26651767, 0.26841034, 0.2567373, 0.30387976, 0.56112034, 0.29374632, 0.25885569, 0.28886394]
	[0.31244987, 0.31077239, 0.31301815, 0.41503845, 0.3028841, 0.51198333, 0.3037852, 0.30417011, 0.51547247, 0.42618875, 0.32133741, 0.4193918, 0.51342593, 0.31474446, 0.30638533, 0.51002859, 0.30073147, 0.29973482, 0.31150254, 0.30698193, 0.51103827, 0.30185637, 0.30465635, 0.30303136]
3	[0.26161177, 0.25226821, 0.25437094, 0.34486372, 0.2705699, 0.43657566, 0.26903233, 0.25832285, 0.43081316, 0.32776564, 0.24156009, 0.34469634, 0.44530685, 0.24884779, 0.26874264, 0.45082217, 0.28355228, 0.27990753, 0.253606, 0.25953102, 0.41827727, 0.27919511, 0.27423923, 0.27572997]
	[0.23222034, 0.27248019, 0.24092578, 0.36419345, 0.30512811, 0.50833465, 0.29578085, 0.31813115, 0.44599348, 0.34194407, 0.24319275, 0.32755039, 0.44986703, 0.24674675, 0.24574968, 0.45980473, 0.27066316, 0.28070829, 0.27058622, 0.30784035, 0.56616587, 0.29369253, 0.25781044, 0.28670078]

Perulangan ke-	Nilai Centroid
	[0.3149749 , 0.31400619, 0.31708373, 0.41480575, 0.30273798, 0.50914718, 0.30440084, 0.30437196, 0.52053367, 0.42357177, 0.32135729, 0.42004006, 0.51524462, 0.31690492, 0.31050617, 0.51180349, 0.30186205, 0.30090137, 0.31345884, 0.30719882, 0.50369157, 0.30174998, 0.30672634, 0.30394261]
4	[[0.26420189,0.25298939, 0.25695867, 0.34573227, 0.27089965, 0.43625981, 0.26790635, 0.2580894 , 0.43002227, 0.32609576, 0.24185571, 0.3441424, 0.44418825, 0.24909853, 0.27027966, 0.45041821, 0.28336086, 0.2788174, 0.25475863, 0.25949348, 0.41431394, 0.28033302,0.27454812,0.27639091] [0.23082128, 0.27290041, 0.23134157, 0.36577895, 0.3030238 , 0.49783379, 0.29357186, 0.30629359, 0.44280912, 0.35449108, 0.24950863, 0.33696547, 0.45491174, 0.24766591, 0.24506258, 0.46250322, 0.27236695, 0.27444482, 0.26460601, 0.30463524, 0.56399964, 0.29078804, 0.25910998, 0.28068124] [0.31587032, 0.31394394, 0.31918146, 0.41461428, 0.30240577, 0.51016158, 0.30503428, 0.30556275, 0.52243394, 0.42299465, 0.32123767, 0.42023316, 0.51574624, 0.31776537, 0.31134678, 0.51215182, 0.3023224 , 0.30288673, 0.31449665, 0.30704901, 0.50252657, 0.30169525, 0.30737874, 0.30489497]

Selain adanya pergeseran *centroid* dalam proses *looping* tersebut juga terdapat pergeseran data dalam setiap *cluster* yang dapat dilihat pada Tabel IV.

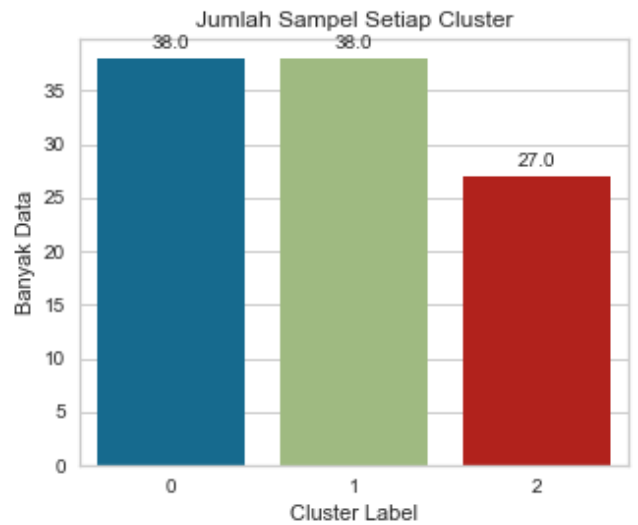
TABEL IV  
LABEL CLUSTER SETIAP LOOPING

Perulangan ke-	Label Klaster
1	[2, 0, 1, 2, 2, 2, 1, 1, 2, 0, 1, 0, 2, 2, 0, 2, 2, 0, 2, 2, 2, 0, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 0, 1, 0, 2, 0, 2, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 2, 2, 0, 0, 0, 2, 2, 0, 2, 2, 2, 0, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 0, 2, 0, 1, 2, 2, 2]
2	[2, 0, 1, 1, 2, 2, 2, 1, 1, 2, 0, 1, 0, 2, 2, 0, 2, 2, 0, 2, 2, 1, 2, 0, 2, 0, 2, 2, 1, 2, 2, 2, 2, 2, 1, 1, 0, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 0, 0, 0, 0, 2, 2, 0, 2, 2, 2, 0, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 0, 2, 0, 1, 2, 2, 2]
3	[2, 0, 1, 1, 2, 1, 2, 1, 1, 2, 0, 1, 0, 2, 2, 0, 2, 2, 0, 2, 2, 1, 2, 0, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 1, 0, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 0, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2, 2,

	0, 2, 2, 2, 0, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 0, 2, 0, 1, 2, 2, 2]
4	[2, 0, 1, 1, 2, 1, 2, 1, 1, 2, 0, 1, 0, 2, 2, 0, 2, 2, 0, 2, 2, 1, 2, 1, 2, 0, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 1, 0, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2, 0, 2, 2, 2, 0, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 0, 2, 0, 1, 2, 2, 2]

### 5. Hasil Clustering

Setelah semua tahapan berhasil dilakukan, *centroid* atau titik pusat *cluster* tidak ada yang bergeser, dan data tidak ada yang bergeser berubah *cluster* lagi, maka didapatkan anggota masing-masing *cluster* seperti diagram pada Gambar 7.



Pada diagram di atas dapat dilihat bahwa data yang termasuk ke dalam label klaster 0 sebanyak 38 data, label klaster 1 sebanyak 38 data, dan label klaster 2 sebanyak 27 data.

### E. Pengukuran Kinerja

Pengukuran kinerja dilakukan dengan menghitung *Error* pada data setelah proses *clustering* dalam setiap metode menggunakan rumus (13). Perbandingan nilai *error* pada setiap metode dapat dilihat pada tabel berikut ini:

TABEL V  
NILAI ERROR PADA DATA SETELAH CLUSTERING

Metode	Error Terendah	Error Tertinggi	Rata-Rata Error
TOPSIS – Kmeans	0.0	0.10583354	0.00886228 43
Kmeans	1.44462956 31 e+19	1.74426552 e+32	2.78629632 78e+30



Berdasarkan tabel di atas, dapat dilihat bahwa nilai *error* pada data setelah proses *clustering* dengan menggunakan kombinasi metode TOPSIS dan algoritma *K-means* lebih rendah daripada nilai *error* pada algoritma *K-means* saja, dengan nilai rata-rata *error* pada data yaitu 0.0088622843.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian mengenai *clustering* bidang keilmuan dengan menggunakan kombinasi metode TOPSIS dan *K-means* yang mengacu pada rumusan masalah, pengujian, serta analisis data hasil pengujian kinerja, dapat disimpulkan beberapa hal sebagai berikut :

1. Penjumlahan terbobot dengan menggunakan metode TOPSIS memiliki performa yang baik dalam kombinasi dengan algoritma *K-means* untuk menentukan *cluster* berdasarkan mata pelajaran yang dikuasai oleh siswa TAPAS AI Manshuroh Jombang.
2. Kinerja algoritma *K-means* dalam *clustering* bidang keilmuan yang berkombinasi dengan metode TOPSIS dinilai cukup optimal. Hal itu diindikasikan dengan nilai *error* pada kombinasi metode TOPSIS dan algoritma *K-means* memiliki rata-rata *error* lebih rendah dari nilai *error* pada *K-means*.

#### V. SARAN

Berdasarkan hasil dari penelitian ini, dapat ditarik saran untuk kepentingan penelitian dan pengembangan berikutnya, yaitu melakukan optimalisasi perhitungan supaya lebih akurat dalam menentukan *cluster* yang sesuai dengan mata pelajaran yang dikuasai oleh masing-masing siswa.

#### REFERENSI

- [1] D. Imanudin and A. Imron, "Hubungan antara Orang Tua dan Lembaga Pendidikan Pendidikan Anak Usia Dini Hubungan antara Orang Tua dan Lembaga Pendidikan Anak Usia Dini," *Paradigm*, vol. 02, no. 02, 2014.
- [2] N. Husna, F. Hanum, and M. F. Azrial, "Pengelompokan Produk Kemasan yang Harus Dihindari Penderita Diabetes Menggunakan Algoritma K-Means Clustering," *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, vol. 4, no. 1, pp. 167–174, Sep. 2019, doi: 10.30743/infotekjar.v4i1.1484.
- [3] M. Anggara, H. Sujiani, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 1, 2016.
- [4] A. Perdana Windarto, "IMPLEMENTASI METODE TOPSIS DAN SAW DALAM MEMBERIKAN REWARD PELANGGAN," *Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 04, 2017.
- [5] D. Hariyanto, R. Malani, and B. Suprpty, "Pengelompokan Bidang Keilmuan Di Teknologi Informasi Dengan Metode K-Means Dan Optimasi Simple Additive Weighting (Saw) Dalam Penentuan Kesesuaian Terhadap Keilmuan," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, 2018.
- [6] S. Sakur, M. Silangen, and D. Tuwohingide, "Penerapan Algoritme K-Means Cluster dan Metode TOPSIS pada Pemilihan Mahasiswa Kunjungan Industri," *urnal Ilmiah Teknik Informatika dan Sistem Informasi*, Dec. 2022.
- [7] J. Wira, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, 1.4. 2020.
- [8] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," vol. 15, no. 2, 2021.
- [9] D. M. Candrasari, A. Syukur, and M. A. Soeleman, "Penentuan Prioritas Penerima Dana Bantuan Operasional Pendidikan Lembaga Pendidikan Anak Usia Dini dengan Metode KNN, TOPSIS dan K-Means," *Jurnal Teknologi Informasi-CyberKU*, vol. 15, Jul. 2019, [Online]. Available: <http://research>.
- [10] W. Yustanti, N. Rahmawati, and Y. Yamasari, "Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 di Propinsi Jawa Timur dengan Metode K-Means," *Journal Information Engineering and Educational Technology*, vol. 04, 2020.
- [11] D. Pribadi, R. Saputra, J. Hudin, and Gunawan, *Sistem Pendukung Keputusan*. Yogyakarta: GRAHA ILMU, 2020.