

OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON

Andre Winarta¹⁾, Wahyu Joni Kurniawan²⁾

^{1,2)}Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia
Jl. Ahmad Yani No, 78-88, Pekanbaru, Riau
E-mail :andre.winarta@student.pelitaindonesia.ac.id¹⁾,
wahyu.jonikurniawan@lecturer.pelitaindonesia.ac.id²⁾

ABSTRACT

As we know, drugs are very illegal in Indonesia because drug abuse can have very dangerous effects. Drugs are substances that have side effects such as hallucinations, reduced awareness and arousal of the user. Technological advances continue to change over time so that information needs are very much needed in life. Currently, data on drug users is very extensive, so that adequate information presentation techniques are needed so that the information received is very accurate and in accordance with the user's needs. Therefore, it is necessary to carry out a data mining process on drug user data to obtain useful information for users. This study aims to prove Elbow's performance to produce optimal clusters of drug user data using the K-Means algorithm as a data grouping method. Cluster optimization is obtained from the Elbow method, which is executed with Google Collaboratory using the Python programming language. The test results show that the Elbow method works very well in producing the optimal cluster, which is found at $k = 3$ with the SSE difference value of 1257.862 with k test = 5.

Keywords: Drug, K-Means, SSE, Elbow, Python

ABSTRAK

Seperti yang kita ketahui bahwa narkoba sangatlah ilegal di Indonesia dikarenakan penyalahgunaan narkoba dapat menimbulkan efek yang sangat berbahaya. Narkoba adalah zat yang memiliki efek samping seperti halusinasi, berkurangnya kesadaran, dan daya rangsang si pengguna. Kemajuan teknologi terus berubah seiring waktu sehingga kebutuhan informasi sangat dibutuhkan dalam kehidupan. Saat ini data mengenai pengguna narkoba sudah sangat luas sehingga teknik penyajian informasi yang memadai sangatlah dibutuhkan supaya informasi yang diterima sangat akurat dan sesuai dengan kebutuhan si pengguna. Maka dari itu, perlu dilakukan proses data mining pada data pengguna narkoba untuk mendapatkan informasi yang bermanfaat bagi para pengguna. Penelitian ini bertujuan untuk membuktikan kinerja *Elbow* untuk menghasilkan *cluster* yang optimal pada data pengguna narkoba menggunakan algoritma *K-Means* sebagai metode pengelompokan data. Optimasi *cluster* diperoleh dari metode *Elbow* yang dieksekusi dengan *Google Colaboratory* yang menggunakan bahasa pemrograman *Python*. Hasil pengujian memperlihatkan bahwa metode *Elbow* bekerja dengan sangat baik dalam menghasilkan *cluster* yang optimal yaitu terdapat pada $k=3$ dengan nilai selisih SSE adalah 1257,862 dengan k uji=5.

Kata kunci: Narkoba, K-Means, SSE, Elbow, Python

1. PENDAHULUAN

Seperti yang kita ketahui bahwa narkoba sangatlah ilegal di Indonesia dikarenakan penyalahgunaan narkoba dapat menimbulkan efek yang sangat berbahaya. Narkoba adalah zat yang memiliki efek samping seperti halusinasi, berkurangnya kesadaran, dan daya rangsang si pengguna. Narkoba atau Narkotika berasal dari bahasa Yunani yaitu *narke* atau *narkam* yang berarti terbius sehingga tidak merasakan apa-apa [1]. Pemakaian narkotika dapat dimanfaatkan dengan syarat harus sesuai dengan aturan-aturan yang sudah ditetapkan.

Kemajuan teknologi terus berubah seiring waktu dimana teknologi semakin lama semakin canggih kedepannya. Dengan adanya teknologi ini, kebutuhan informasi sangat dibutuhkan dalam kehidupan. Namun nyatanya kebutuhan informasi saat ini masih belum memadai, dan terkadang dari informasi tersebut, data masih harus digali lagi untuk mendapatkan informasi yang akurat.

Saat ini data mengenai penggunaan narkoba sudah sangat luas sehingga teknik penyajian informasi yang memadai sangatlah dibutuhkan supaya informasi yang diterima sangat akurat dan sesuai dengan kebutuhan si pengguna. Maka dari itu, sangat penting untuk membangun sebuah algoritma *data mining* yang sangat efisien dan akurat terhadap jumlah data yang sangat besar.

Data mining merupakan proses dengan kecerdasan buatan, teknik statistik, matematika, dan *machine learning* untuk mengidentifikasi dan mengekstraksi dan suatu informasi yang bermanfaat dan pengetahuan terkait dari berbagai *database* yang besar [2] yang dimana salah satu metode dari *data mining* tersebut adalah *Clustering*.

Clustering adalah suatu metode pengelompokan data. *Clustering* merupakan suatu metode pengelompokan data kedalam beberapa *cluster* (kelompok) dimana memiliki tingkat kemiripan yang

maksimum antar data didalam satu *cluster* dan tingkat kemiripan yang minimum terhadap antar *cluster*. Salah satu teknik *clustering* adalah algoritma *K-Means*.

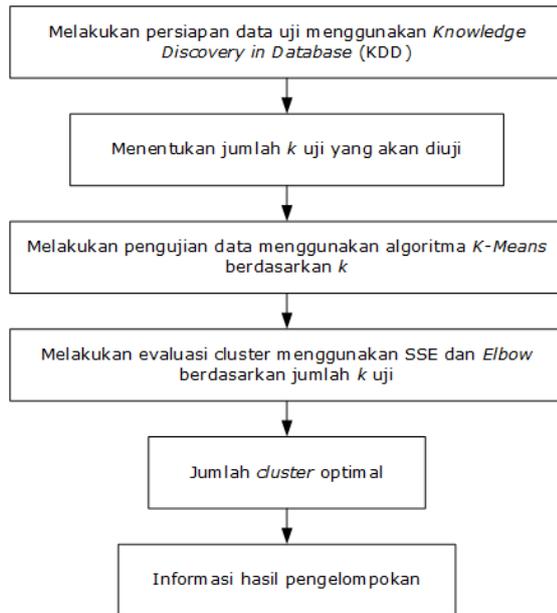
Untuk memperkuat latar belakang permasalahan, penulis memasukkan jurnal yang berkaitan dengan algoritmanya yang diteliti oleh *Muningsih dan Kiswati dengan judul "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan"*. Dalam penelitian tersebut menjelaskan bahwa pengolahan data yang dilakukan dengan menggunakan metode *K-Means* dan optimasi metode elbow dengan mengetahui nilai SSE (Sum of Square Error) dihasilkan 3 kelompok pelanggan yang memiliki nilai maksimal atau terbaik dengan hasil selisih SSE sebesar 3124559,296.

Dengan adanya penelitian ini, penulis berharap hasil *cluster* dengan menggunakan metode *K-Means* dan optimasi jumlah *cluster* dengan menggunakan metode *elbow* dapat menghasilkan *cluster* yang optimal dan juga dapat menghasilkan informasi-informasi yang bermanfaat

2. METODOLOGI PENELITIAN

Dari penelitian yang dilakukan oleh *Muningsih dan Kiswati dengan judul "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan"* kesimpulan dari penelitian ini adalah melakukan optimasi pada *cluster K-Means* untuk menghasilkan *cluster* yang optimal dan informasi-informasi yang bermanfaat.

Untuk memudahkan penelitian, disusun kerangka penelitian pada Gambar 1.

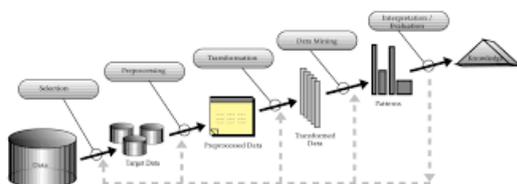


Gambar 1. Kerangka Penelitian

2.1 Melakukan Persiapan Data Uji

Penelitian diawali dengan mempersiapkan data untuk diuji. Persiapan data mengikuti langkah yang terdapat dalam *Knowledge Discovery in Database (KDD)*.

Menurut Larose [3], ada beberapa tahapan pada proses *data mining* diawali dari penyeleksian data, proses *cleaning* data, proses transformasi, proses *data mining* atau proses mencari pola atau informasi dari sebuah data terpilih dan tahap terakhir adalah tahap interpretasi dan evaluasi yang menghasilkan informasi-informasi baru yang bermanfaat.



Gambar 2. Tahapan Data Mining

Sumber : (Jollyta et al, 2020) [4]

Tahapan yang dilakukan pada proses *data mining* diawali dari *data selection*, *pre-processing data*, *transformation data*, *data mining*,

interpretation / evaluation. Secara detail dijelaskan sebagai berikut:

1. Data Selection

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (*Knowledge Discovery in Database*). Data hasil seleksi akan digunakan dalam proses *data mining*.

2. Pre-processing / Cleaning

Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Proses *transformation* ini adalah proses dimana data yang telah dipilih akan diubah kedalam bentuk dimana data bisa diproses dalam *data mining*

4. Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. Interpretation / Evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.2 Menentukan Jumlah K Uji

Setelah data uji melewati proses tersebut dalam tahapan KDD, data siap dipakai ke tahap selanjutnya yaitu menentukan jumlah *k* uji yang akan diuji.

2.3 Pengujian Algoritma K-Means

Setelah jumlah *k* uji ditentukan, selanjutnya data diuji menggunakan algoritma *K-Means* berdasarkan nilai *k*.

Algoritma *K-Means* adalah suatu metode penganalisaan data yang melakukan proses pengelompokan data dengan sistem partisi. Menurut Yahya dan Mahpuz [5], Algoritma *K-Means*

merupakan model yang menggunakan *centroid* untuk membuat *cluster*. *Centroid* adalah titik tengah dari suatu *cluster*.

Secara umum algoritma K-Means memiliki langkah-langkah seperti berikut:

1. Tentukan *k* sebagai jumlah *cluster* yang akan dibentuk
2. Tentukan *k Centroid* (titik pusat *cluster*) awal secara acak.

Adapun rumus yang digunakan untuk menghitung *centroid cluster* ke-*i* berikutnya sebagai berikut:

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

Dimana :

v : *centroid* pada *cluster*

x_i : objek ke-*i*

n : banyaknya objek / jumlah objek yang menjadi anggota *cluster*

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*.

Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan Euclidean Distance.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3, \dots, n \quad (2)$$

Dimana :

x_i : objek *x* ke-*i*

y_i : data *y* ke-*i*

n : banyaknya objek

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1)
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama [6]

2.4 Evaluasi Cluster Menggunakan SSE dan Elbow

Setelah melewati proses algoritma *K-Means*, hasil pengelompokan data tiap *k* akan divalidasi menggunakan SSE dan *Elbow*.

Menurut Putu [7], Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam

menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster (*k*) yang akan membentuk siku pada suatu titik.

Nilai *k* pada kombinasi siku dengan K-Means adalah grafik hubungan cluster dengan penurunan error [8]. Jumlah cluster *k* yang dihasilkan dari pengujian dengan K-Means dievaluasi menggunakan teknik SSE [8]. SSE (Sum of Square Error) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya [9]. Untuk menghitung SSE dapat menggunakan rumus sebagai berikut :

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - y_k\|_2^2 \quad (3)$$

2.5 Jumlah Cluster Optimal

Nilai SSE yang sudah didapat akan diuji menggunakan metode *Elbow* untuk menentukan jumlah *cluster* yang optimal.

2.6 Informasi Hasil Pengelompokan

Setelah didapatkan jumlah *cluster* yang optimal, diperoleh informasi hasil pengelompokan berdasarkan *cluster* yang optimal

3. HASIL DAN PEMBAHASAN

3.1 Data Uji (Training)

Data Pengguna Narkoba merupakan data set yang terdiri dari 1000 data yang sudah dikonversi ke dalam bentuk numerik. Data ditampilkan pada Tabel 1.

Tabel 1. Data Pengguna Narkoba

No	Rating	Effectiveness	Side Effects
1	1	5	2
2	10	5	5
3	3	2	4
4	2	2	2
5	1	1	2
6	9	5	4
7	10	4	5
8	10	5	5
9	1	1	1

10	7	4	3
...			
1000	1	2	2

3.2 Hasil Pengelompokan K-Means

Data pengguna narkoba diujikan melalui algoritma *K-Means* dengan jumlah *cluster* uji mulai dari $k=2$ hingga $k=6$. Jumlah anggota tiap pengelompokan ditampilkan pada Tabel 2 berikut.

Tabel 2. Pengelompokan K-Means

Jumlah k	Jumlah Anggota Cluster
2	cluster 1(262 data), cluster 2(738 data)
3	cluster 1(186 data), cluster 2(579 data), cluster 3(235 data)
4	cluster 1(107 data), cluster 2(544 data), cluster 3(234 data), cluster 4(115 data)
5	cluster 1(107 data), cluster 2(257 data), cluster 3(184 data), cluster 4(114 data), cluster 5(338 data)
6	cluster 1(82 data), cluster 2(257 data), cluster 3(158 data), cluster 4(92 data), cluster 5(338 data), cluster 6(73 data)

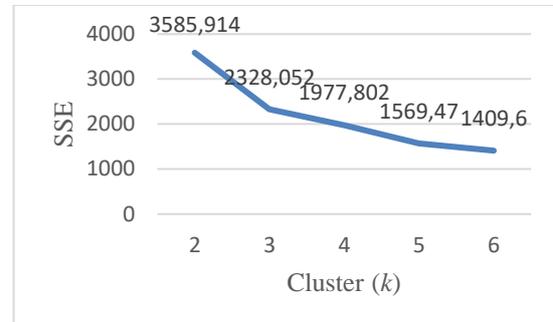
3.3 Hasil Evaluasi SSE dan Elbow

Dari hasil proses perhitungan SSE terhadap 1000 data pengguna narkoba maka hasil yang mengalami penurunan yang paling besar adalah pada $k=3$ dengan nilai selisih sebesar 1257,862. Ini dapat dilihat pada Tabel 3 dan Gambar 3.1.

Tabel 3. Hasil SSE Pada Tiap Cluster

Cluster(k)	Hasil SSE	Selisih
2	3585,914	-

3	2328,052	1257,862
4	1977,802	350,25
5	1569,47	408,332
6	1409,6	159,87



Gambar 3. Grafik SSE

3.4 Informasi Pengelompokan

Berdasarkan Tabel 3, nilai selisih tertinggi terdapat pada $k=3$ yakni 1257,862. Nilai ini menunjukkan bahwa *cluster* optimal terdapat pada $k=3$. Artinya, selain *cluster* yang optimal, informasi yang terkandung didalamnya juga merupakan informasi yang terbaik.

Dilihat dari hasil pengelompokan $k=3$ yang ditampilkan mulai dari cluster 1 hingga cluster 3, memiliki jumlah anggota masing-masing cluster adalah 186, 579, 235. Informasi yang diperoleh adalah sebagai berikut.

Tabel 4. Informasi Berdasarkan Cluster Optimal

Cluster	Informasi
1(186)	Kelompok ini merupakan penggunaan narkoba dengan rating yang kurang, kurang efektif dan memiliki efek samping yang cukup tinggi
2(579)	Kelompok ini merupakan penggunaan narkoba dengan rating yang sangat bagus dengan efektivitas yang tinggi dan memiliki efek samping yang rendah
3(235)	Kelompok ini merupakan penggunaan narkoba dengan

rating yang cukup bagus, efektivitas yang cukup tinggi dan efek samping yang tidak terlalu tinggi

3.5 Implementasi Pada Python

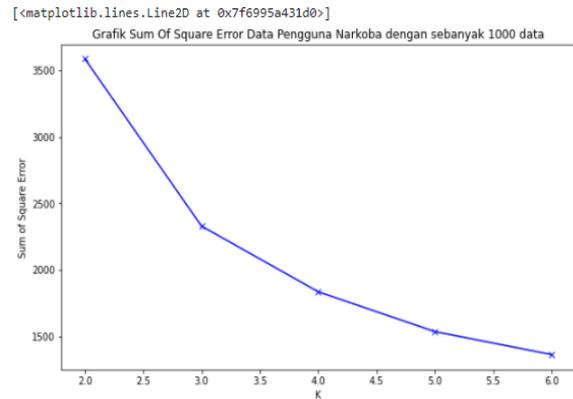
Hasil eksekusi Python untuk pengelompokan menggunakan *K-Means* adalah sebagai berikut.

Tabel 5. Hasil *K-Means* K=2

```
Masukkan jumlah k uji : 5
K-Means dengan k = 2
Centroid 0
[8.58152174 4.41711957 4.11005435]
RATING EFFECTIVENESS (NUMERIK) SIDE EFFECTS (NUMERIK) Cluster
NO
2 10 5 5 0
6 9 5 4 0
7 10 4 5 0
8 10 5 5 0
10 7 4 3 0
... ..
995 8 4 4 0
996 8 4 5 0
997 9 4 4 0
998 7 4 3 0
999 8 5 4 0
[736 rows x 4 columns]

Centroid 1
[2.70075758 2.63636364 2.55681818]
RATING EFFECTIVENESS (NUMERIK) SIDE EFFECTS (NUMERIK) Cluster
NO
1 1 5 2
1 3 2 4
1 2 2 2
5 1 1 2
1 9 1 1
1 ... ..
... ..
973 5 3 2
1 985 2 5 3
1 986 5 5 4
1 988 3 3 2
1 1000 1 2 2
1
[264 rows x 4 columns]
```

Berikut grafik SSE yang dieksekusi dengan Python.



Gambar 3. Grafik SSE Pada Python

4 KESIMPULAN

Berdasarkan pembahasan yang telah dilakukan, *K-Means* berhasil membentuk pengelompokan dengan anggota *cluster* yang tepat dan dioperasikan menggunakan pemrograman Python pada data pengguna narkoba. Melalui eksekusi Python, informasi yang dihasilkan sesuai dengan keadaan yang ada karena muncul dari *cluster* yang optimal hasil perhitungan SSE dengan nilai selisih tertinggi berada pada $k=3$. Namun pada beberapa *cluster* terdapat data yang dianggap tidak berada pada *cluster* yang tepat. Hal ini sangat mungkin terjadi karena penentuan titik pusat *cluster* pengujian awal yang ditentukan sendiri oleh Python dan tidak dapat diubah sehingga mempengaruhi pula perhitungan yang ada.

DAFTAR PUSTAKA

[1] F. Eleanora, “BAHAYA PENYALAHGUNAAN NARKOBA SERTA USAHA PENCEGAHAN DAN PENANGGULANGANNYA (Suatu Tinjauan Teoritis),” *J. Huk.*, vol. 25, no. 1, pp. 439–452, 2011.

[2] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[3] Larose, “Discovering Knowledge in

- Data: An Introduction to Data Mining,” pp. 129–240, 2005.
- [4] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data Mining Dan Penerapan*, Pertama. Yogyakarta: Deepublish, 2020.
- [5] Yahya and Mahpuz, “Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat,” *J. Inform. dan Teknol.*, vol. 2, no. 2, pp. 109–118, 2019.
- [6] Ediyanto, N. Mara, and N. S. Intisari, “Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis,” *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- [7] N. Putu, E. Merliana, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” pp. 978–979.
- [8] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, “Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.
- [9] A. T. Rahman, Wiranto, and A. Rini, “Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama),” *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017, doi: 10.20961/ITS.V6I1.11296.