

# Anomali Data Mining Menggunakan Metode K-Means Dalam Penilaian Mahasiswa Terhadap Pelayanan Prodi

Deti Karmanita<sup>1</sup>, Billy Hendrik<sup>2</sup>

<sup>1</sup> Mahasiswa Fakultas Ilmu Komputer Universitas Putra Indonesia YPTK Padang. E-mail: [detikarmanita07@gmail.com](mailto:detikarmanita07@gmail.com)

<sup>2</sup> Dosen Tetap Fakultas Ilmu Komputer Universitas Putra Indonesia YPTK Padang E-mail: [billy\\_hendrik@upiptyk.ac.id](mailto:billy_hendrik@upiptyk.ac.id)

Jl. Raya Lubuk Begalung, Lubuk Begalung Nan XX, Kec. Lubuk Begalung, Kota Padang, Sumatera Barat 25145

(Received: Mei 2023, Revised : Agustus 2023, Accepted : Oktober 2023)

**Abstract:** Cluster analysis is a data mining technique that aims to identify a group of objects that have the same characteristics. The number of groups that can be identified depends on the amount of data and the type of object, so that data problems arise when there is a change to a number of redundant data, but not all of it is changed where the data above is repeatedly made into one table with a different code as the primary key and there are anomalies Insertion, so K-means is one method of clustering data which is divided into the form of one or more clusters/groups that have the same characteristics. Student data clustering uses the k-means method, consisting of student assessments. This study uses student assessment data. Then it was concluded that the assessment group was based on reliability aspects: the ability of lecturers, education staff and administrators to provide services, responsiveness aspects: the willingness of lecturers, education staff and administrators to help students and provide services quickly, aspects of certainty (assurance): the ability of lecturers, staff and administrators to give confidence to students that the services provided are in accordance with the provisions, aspects of empathy (empathy): the willingness/concern of lecturers, staff and managers to give attention to students, tangibles aspects: students' assessment of the adequacy, accessibility, quality of facilities and infrastructure from the grouping results based on reliability, responsiveness, assurance and empathy data.

**Keyword :** Data Mining, Anomali, K-Mens

**Intisari:** Analisis cluster merupakan teknik data mining yang bertujuan untuk mengidentifikasi suatu kelompok dari objek yang memiliki karakteristik yang sama. Jumlah kelompok yang dapat diidentifikasi tergantung pada sejumlah data dan jenis dari objeknya. Sehingga data masalah yang timbul saat terjadi perubahan terhadap sejumlah data yang mubazir, tetapi tidak seluruhnya diubah dimana data diatas berulang dijadikan dalam satu tabel dengan kode yang berbeda sebagai primary keynya dan terdapat Anomali Penyisipan, sehingga K-means adalah salah satu metode clustering data yang dibagi kedalam bentuk satu atau lebih cluster/kelompok yang memiliki karakteristik sama. Clustering data mahasiswa menggunakan metode k-means, terdiri dari penilaian mahasiswa. Penelitian ini menggunakan data penilaian mahasiswa. Kemudian diperoleh kesimpulan bahwa kelompok penilaian berdasarkan aspek keandalan

(reliability) : kemampuan dosen, tenaga kependidikan dan pengelola dalam memberikan pelayanan, aspek daya tanggap (responsiveness) : kemauan dosen, tenaga kependidikan dan pengelola dalam membantu mahasiswa dan memberikan jasa dengan cepat, aspek kepastian (assurance) : kemampuan dosen, tendik dan pengelola untuk memberikan keyakinan kepada mahasiswa bahwa pelayanan yang diberikan telah sesuai dengan ketentuan, aspek empati (empathy) : kesediaan/kepedulian dosen, tendik dan pengelola untuk memberi perhatian kepada mahasiswa, aspek tangibles : penilaian mahasiswa terhadap kecukupan, aksesibilitas, kualitas sarana dan prasarana dari hasil pengelompokan berdasarkan data reability, responsiveness, assurance dan empati.

**Keyword :** Data Mining, Anomali, K-Mens

## I. PENDAHULUAN

Perguruan tinggi sebagai institusi pendidikan telah memiliki data akademik dan administrasi dalam jumlah yang sangat besar, namun hanya sebagian kecil data tersebut dimanfaatkan (khususnya dalam penyusunan evaluasi diri). Kuesioner kepuasan mahasiswa terhadap layanan pendidikan di pada setiap awal tahun akademik melakukan kepuasan mahasiswa terhadap layanan pendidikan di universitas dehasen bengkulu dimana data mahasiswa baru yang dihimpun pada saat pendaftaran dan/atau registrasi berupa nilai kepuasan penilaian mahasiswa terhadap kecukupan, aksesibilitas, kualitas sarana dan prasarana dari hasil pengelompokan berdasarkan data reability, responsiveness, assurance dan empati berupa anek yang masuk perguruan tinggi. Jumlah data yang terakumulasi dari tahun ke tahun perlu dilakukan analisis untuk dapat membuka peluang dihasilkannya informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen

perguruan tinggi. Untuk menarik minat calon mahasiswa yang lebih banyak, diperlukan strategi promosi yang tepat, objek penelitian bersumber dari data mahasiswa yang telah melakukan registrasi di universitas dehasen bengkulu. Atribut data yang akan digunakan adalah kuesioner penilaian mahasiswa terhadap kecukupan, aksesibilitas, kualitas sarana dan prasarana dari hasil pengelompokan berdasarkan data reability, responsiveness, assurance dan empati, hasil yang diperoleh dari penelitian ini diharapkan dapat menampilkan, keterkaitan antara hasil analisis tersebut diharapkan dapat membantu pihak admisi perguruan tinggi dalam menyusun strategi promosi program studi yang menjadi target. Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah record dan jumlah atribut yang cukup banyak) tidak dapat dilakukan dengan mudah. Teknologi data mining merupakan salah satu alat bantu untuk penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah record dan jumlah atribut yang cukup banyak) tidak dapat dilakukan dengan mudah. Teknologi data mining merupakan salah satu alat bantu untuk penggalan data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak digunakan pada lingkungan aplikasi bisnis seperti perbankan, provider telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain, clustering merupakan salah satu teknik data mining yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap cluster akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya sehingga anomali adalah proses pada basis data yang memberikan efek

samping yang tidak diharapkan, misalnya menyebabkan ketidakkonsistenan data atau membuat suatu data menjadi hilang. Data anomali penduduk merupakan data penduduk yang tidak lazim dan belum diketahui kebenarannya.

## II. TINJAUAN PUSTAKA

Penelitian ini menggunakan teknik klasifikasi dengan K-Means untuk mengelompokkan data kategorikal hingga menghasilkan kluster yang lebih stabil. Selanjutnya melakukan analisis lebih mendalam terhadap hasil klusterisasi tersebut untuk membantu bagian admisi dalam menentukan kebijakan di lingkungan Universitas Dehasen Bengkulu. Analisis cluster merupakan salah satu teknik data mining yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek lainnya, sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek yang berada pada kelompok yang berbeda. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data objek. Tujuan dari pengelompokan sekumpulan data objek ke dalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan. Model yang diambil diasumsikan bahwa data yang dapat digunakan adalah data yang berupa data interval, frekuensi dan biner. Set data objek harus mempunyai variabel dengan tipe yang sejenis tidak campur antara tipe yang satu dengan lainnya. Algoritma K-Means dimulai dengan pemilihan secara acak K, yang merupakan banyaknya kluster yang ingin dibentuk. Kemudian ditetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari

kluster atau biasa disebut dengan centroid/mean. Selanjutnya dihitung jarak dari setiap data yang ada terhadap masing-masing centroid menggunakan rumus yang sudah disediakan hingga diketemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasi setiap data berdasarkan kedekatannya dengan centroid.

**III. METODOLOGI PENELITIAN**

Metode clustering data yang dibagi kedalam bentuk satu atau lebih cluster/kelompok yang memiliki karakteristik sama. Clustering data mahasiswa menggunakan metode k-means, terdiri dari penilaian mahasiswa.

**IV. HASIL DAN PEMBAHASAN**

Anomali merupakan proses pada basis data yang memberikan efek samping yang tidak diharapkan, Anomali ini terjadi karena struktur tabel yang kita buat memiliki masalah. Masalah ini akan meningkatkan kompleksitas data dan menyulitkan dalam pemodifikasian data itu sendiri. Sehingga harus dilakukan cara-cara yang disebut dengan normalisasi.

No.	Pertanyaan	Total	Rata-Rata
<b>A.</b> Aspek Seseorang (Individuality) : Kemampuan Dosen, Tenaga Kependidikan dan Pengelola dalam memberikan pelayanan			
1	Dosen menerima dan menanggapi materi yang disampaikan	170	3,8
2	Dosen menerima dan menanggapi sistem belajar yang dikembangkan.	168	3,7
3	Staf fakultas memberikan layanan prima.	171	3,8
4	Staf prodi memberikan layanan prima.	166	3,7
5	Dekan memberikan layanan prima.	168	3,8
6	Rekan Program Studi memberikan layanan prima.	166	3,7
<b>B.</b> Aspek Gaya Tanggap (Responsiveness) : Kemauan Dosen, tenaga kependidikan dan pengelola dalam membantu mahasiswa dan memberikan jasa dengan cepat.			
1	Dosen PA cepat tanggap dalam membantu kepentingan mahasiswa.	180	4,0
2	Dosen Pembina yang terampil/TA cepat tanggap membantu mahasiswa dalam proses pembelajaran.	178	3,9
3	Staf fakultas cepat tanggap dalam melayani kebutuhan mahasiswa	178	3,9
4	Staf prodi cepat tanggap dalam melayani kebutuhan mahasiswa	166	3,7
5	Dekan cepat dan tanggap dalam melayani kepentingan mahasiswa	181	4,0
6	Rekan Program Studi cepat dan tanggap dalam melayani kepentingan mahasiswa	180	4,0
<b>C.</b> Aspek Kepastian (Assurance) : Kemampuan dosen, tenik dan pengelola untuk memberikan keyakinan kepada mahasiswa bahwa pelayanan yang diberikan telah sesuai dengan ketentuan.			
1	Dosen memberikan penilaian kepada mahasiswa sesuai dengan standar	170	3,8
2	Dosen mudah untuk dihubungi saat di luar jam kerja.	168	3,7
3	Staf fakultas memberikan pelayanan sesuai dengan standar	170	3,8
4	Staf Prodi memberikan pelayanan sesuai dengan standar	168	3,7

Anomali Peremajaan Adalah masalah yang timbul saat terjadi perubahan terhadap sejumlah data yang mubazir, tetapi tidak seluruhnya diubah dimana data diatas berulang dijadikan dalam satu tabel dengan kode yang berbeda sebagai primary keynya dan terdapat Anomali Penyisipan, Adalah masalah yang timbul saat melakukan penambahan data. Dan data itu adalah Primary Key atau elemen kunci sehingga Anomali Penyisipan, solusinya adalah membuat 2 tabel berbeda, 1 tabel untuk data mahasiswa dan satu lagi tabel untuk penilaian, Untuk dapat menambahkan data penilaian yang baru kita hanya perlu menambahkan data pada tabel diatas Sehingga datanya lebih dinamis dan mencegah dari kolom pada tabel yang tidak terisi, sehingga Anomali Penghapusan, Adalah masalah yang timbul saat melakukan penghapusan data, dan data lain ikut terhapus atau hilang.

1. Analisa Data yang digunakan adalah data diatas
2. Selanjutnya diproses menggunakan algoritma K-Means, dengan urutan langkah sebagai berikut.
  - a. Menentukan koordinat titik tengah setiap cluster,
  - b. Menentukan jarak setiap objek terhadap koordinat titik tengah,
  - c. Mengelompokkan objek-objek tersebut berdasarkan pada jarak minimumnya.
  - d. Tentukan pusat cluster baru
  - e. Apakah ada selisih antara cluster lama dengan cluster baru, Jika masih ada kembali ke langkah-1, jika tidak lanjut kelangkah
  - f. Selesai.
3. Iterasi 1
  - a. Penentuan nilai awal titik tengah  
Langkah pertama, menentukan pusat cluster secara acak pada data awal yang ada di Tabel I. Cluster pertama yang akan digunakan dalam perhitungan adalah C1 (160, 1.8),

artinya Nilai Rerata total dari kuesioner pada pengelompokan pertama adalah 158.1.7 dan penilaian sebesar 1.8. Cluster kedua C2 (180, 1.9) dan cluster ketiga C3 (323, 3.4).

- b. Menghitung jarak objek ke centroid dengan menggunakan rumus jarak Euclid  
Perhitungan jarak penilaian pertama dengan pusat cluster pertama adalah

$$C11 =$$

$$\sqrt{(160 - 158)^2 + (1.8 - 3.4)^2} = 2.56$$

Perhitungan jarak penilaian pertama dengan pusat cluster kedua adalah

$$C12 =$$

$$\sqrt{(172 - 180)^2 + (1.9 - 1.8)^2} = 64.01$$

Perhitungan jarak penilaian kedua dengan pusat cluster ketiga adalah

$$C13 =$$

$$\sqrt{(324 - 323)^2 + (3.4 - 3.4)^2} = 1$$

Perhitungan Yang sama dilakukan untuk data berikutnya, sehingga hasil didapat

$$\text{Penilaian } 1 = C1 = 2.56 \quad C2 = 64.01 \quad C3 = 1$$

Melakukan clustering objek dengan memasukkan setiap objek ke dalam cluster (grup) berdasarkan jarak minimumnya. Suatu data akan menjadi anggota dari suatu cluster (C1, C2 maupun C3) yang memiliki jarak terkecil, dari pusat cluster-nya. Misalnya untuk data pertama, jarak terkecil ada pada cluster yang ke-3 yaitu 1, berarti penilaian pertama akan menjadi anggota pada kelompok cluster yang ke-2. Demikian juga untuk data yang kedua, jarak terkecil ada pada cluster yang ke-3 yaitu 1, berarti penilaian kedua akan menjadi anggota pada kelompok cluster yang ke-3. Posisi cluster setiap penilaian.

- c. Langkah berikutnya menghitung pusat cluster baru

Pusat cluster baru ditentukan berdasarkan pengelompokan anggota dari masing-masing cluster. Berdasarkan Tabel diatas, cluster pertama (C11) dengan parameter Nilai Rerata penilaian. Pusat cluster pertama yang baru dihitung berdasarkan rata-rata koordinat dari penilaian tersebut

- d. Ulangi iterasi yang dimulai dari langkah 2, sampai konvergen sehingga cluster yang baru memiliki nilai yang tetap (tidak mengalami perubahan).

## V. PENUTUP

- a. Simpulan Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma K-Means bisa digunakan untuk mengelompokkan penilaian sehingga data anomali yang tidak sesuai dihapus secara otomatis sehingga tidak ada data yang ganda meliputi: a. Cluster yang pertama memiliki pusat cluster Cluster pertama yang akan digunakan dalam perhitungan adalah C1 (160, 1.8), artinya Nilai Rerata total dari kuesioner pada pengelompokan pertama adalah 158.17 dan penilaian sebesar 1.8. Cluster kedua C2 (180, 1.9) dan cluster ketiga C3 (323, 3.4), Perhitungan Yang sama dilakukan untuk data berikutnya, sehingga hasil didapat Penilaian  $1 = C1 = 2.56$   $C2 = 64.01$   $C3 = 1$
- b. Saran Untuk mengembangkan lebih jauh dari penelitian ini, disarankan untuk mengembangkan lebih lanjut adanya pemetaan kesesuaian antara penialain dikaitkan dengan sistem aplikasi yang ada dikampus.

## DAFTAR PUSTAKA

- [1] A.S. Rosa, M. Shalahuddin. (2013). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*, Informatika. Bandung.
- [2] Dr. Marty Oesman, Yevis. (2010). *Sukses Mengelola Marketing Mix, CRM, Customer Value, dan Customer Dependency*. Bandung: Informasi – Volume 4 No 2 – Oktober 2012, ISSN PRINT 2085-1588, ISSN ONLINE 2385-4614
- [3] Kristanto, Andi. (2010). *Kupas Tuntas PHP & MySQL*. Bandung
- [4] Nugroho, Bunafit. (2014). *Database Relasiona Dengan MySQL*. Jakarta
- [5] Ovi Dyantina, Mira Afrina, Ali Ibrahim. (2012). *Penerapan Customer Relationship Management (CRM) Berbasis Web Pada Toko Yen-Yen*. Jurnal Sistem Informasi – Volume 4 No 2 – Oktober 2012, ISSN PRINT 2085-1588, ISSN ONLINE 2385-4614
- [6] Rahajo, Budi. (2012). *Modul Pemrograman Web (HTML, PHP, & MySql)*. Informatika
- [7] D. Dillenberger, P. Novotny, Q. Zhang, P. Jayachandran, H. Gupta, S. Hans, et al., *Blockchain analytics and artificial intelligence*, IBM J. Res. Dev. 63(2/3) (2019) 5:1–5:14.
- [8] J. Mendling, I. Weber, W.M.P.v.d. Aalst, J.v. Brocke, C. Cabanillas, F. Daniel, S. Debois, C.D. Ciccio, M. Dumas, S. Dustdar, et al., *Blockchains for business process management - challenges and opportunities*, ACM Trans. Manag. Inf. Syst. 9 (1) (2018) 4:1–4:16.
- [9] R. Hobeck, C. Klinkmüller, D. Bandara, I. Weber, *Process mining on Anomali data: A case study of Augur*, in: 19th International Conference, BPM, Rome, Italy, Springer, 2021, pp. 306–323.
- [10] R. Mühlberger, S. Bachhofner, C. Di Ciccio, L. García-Bañuelos, O. Pintado, *Extracting event logs for process mining from data stored on the blockchain*, in: BPM International Workshops, Vienna, Austria, in: LNBIP, vol.362, Springer, 2019, pp. 690–703.
- [11] N.Y. Wirawan, B.N. Yahya, H. Bae, *Incorporating transaction lifecycle information in Anomali process discovery*, in: *Blockchain Technology for IoT Applications*, Springer Singapore, 2021, pp. 155–172.

- [12] L. Moctar-M'Baba, M. Sellami, W. Gaaloul, M.F. Nanne, Blockchain logging for process mining: A systematic review, in: 55th Hawaii International Conference on System Sciences, HICSS, Virtual Event / Maui, Hawaii, USA, January 4-7, ScholarSpace, 2022, pp. 1–10.
  
- [13] IEEE standard for extensible event stream (XES) for achieving interoperability in event logs and event streams, in: IEEE Std 1849-2016, 2016, pp. 1–50, <http://dx.doi.org/10.1109/IEEESTD.2016.7740858>.