

DATA MINING PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN PENERIMA BANTUAN KARTU INDONESIA PINTAR (STUDI KASUS DI UNIVERSITAS MUSI RAWAS)

Siska Aprilia¹, Elmayati², Ahmad Sobri³

¹²³Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau

³Program Studi Informatika, Universitas Bina Insan, Lubuklinggau

e-mail: *¹Siskaaprilias381@gmail.com, ²elmayati@univbinainsan.ac.id,
³ahmadsobri506@gmail.com

Abstrak

Masalah dalam penelitian ini adalah belum meratanya pembagian KIP yang dilakukan oleh Universitas Musi Rawas dan kesulitan mengelompokkan data mahasiswa untuk menentukan prioritas utama dalam penerima bantuan KIP menjadi kendala bagi pihak kampus dalam menghitung setiap data dan menjadikan waktu kurang efektif dan efisien. Metode penelitian yang digunakan adalah metode analisis dengan melakukan pengelompokkan menggunakan *k-means clustering*. Teknik pengumpulan data yang digunakan adalah observasi, wawancara, dokumentasi, dan studi pustaka. Hasil penelitian ini adalah *clustering* menggunakan program Rstudio. *Clustering* dilakukan dengan beberapa proses dan *library* dalam program agar algoritma *clustering* dapat dijalankan. Proses *clustering* dilakukan dengan mempersiapkan dataset berupa data mahasiswa penerima KIP Universitas Musi Rawas Tahun 2022. Kemudian mempersiapkan *library* di dalam program Rstudio sesuai dengan keperluan, seperti *cluster*, *factoextra*, *ggplot2*, dan *tidyverse*. Proses selanjutnya yang dilakukan adalah *import* data dari excel ke Rstudio. *Import* dilakukan dengan mengcopy data dari excel kemudian membentuk kode di dalam Rstudio yang dapat membaca data yang kita copy dengan perintah *read.delim*. Kemudian menampilkan data hasil *import* ke dalam bentuk tabel dan dalam bentuk deskripsi. Setelah data berhasil ditampilkan, langkah selanjutnya adalah melakukan transformasi data. Setelah proses transformasi selesai, langkah selanjutnya yaitu menampilkan data hasil transformasi. Proses selanjutnya yaitu menentukan jumlah *cluster* dengan *k-means*. Kemudian menampilkan hasil *clustering* dalam bentuk grafik dan tabel. Langkah terakhir yang dilakukan adalah melihat deskripsi hasil data *clustering* berdasarkan *cluster*. Nilai rata-rata tertinggi untuk program studi berada pada *cluster* 2 sebesar 6,09. Nilai rata-rata tertinggi untuk jumlah saudara berada pada *cluster* 1 sebesar 3,85. Nilai rata-rata tertinggi untuk penghasilan orang tua berada pada *cluster* 4 sebesar 25000000. Nilai rata-rata tertinggi untuk pekerjaan orang tua berada pada *cluster* 4 sebesar 5. Nilai rata-rata tertinggi untuk status orang tua berada pada *cluster* 3 sebesar 2.

Kata Kunci : KIP, Clustering, Rstudio

Abstract

The problem in this research is the uneven distribution of KIP conducted by MUSI RAWAS UNIVERSITY and the difficulty of grouping student data to determine the top priority of KIP aid recipients which is an obstacle for the campus in calculating each data and making time less effective and efficient. The research method used is the method of analysis by grouping using *k-means clustering*. Data collection techniques used are observation, interviews, documentation, and literature study. The result of this research is clustering using the Rstudio program. Clustering is done with several processes and libraries in the program so that the clustering algorithm can be run. The clustering process is carried out by preparing a dataset in the form of data on students receiving KIP at Musi Rawas University in 2022. Then preparing

the library in the Rstudio program as needed, such as clusters, factoextra, ggplot2, and tidyverse. The next process that is carried out is importing data from excel to Rstudio. Import is done by copying data from excel then forming code in Rstudio that can read the data we copy with the read.delim command. Then displays the imported data in table form and in the form of a description. After the data is successfully displayed, the next step is to transform the data. After the transformation process is complete, the next step is to display the transformed data. The next process is to determine the number of clusters with k-means. Then display the clustering results in the form of graphs and tables. The final step is to look at the description of the results of clustering data based on clusters. The highest average score for study programs is in cluster 2 of 6.09. The highest average value for the number of siblings is in cluster 1 of 3.85. The highest average score for parents' income is in cluster 4 of 25000000. The highest average score for parents' occupation is in cluster 4 of 5. The highest average score for parental status is in cluster 3 of 2.

Keywords : KIP, Clustering, Rstudio

I. PENDAHULUAN

KIP merupakan program pemerintah yang diluncurkan untuk mengatasi masalah yang terjadi karena masih banyak ditemukan kasus siswa yang masih usia sekolah namun putus sekolah karena kesulitan biaya. Persyaratan bagi penerima KIP pada jenjang perguruan tinggi adalah 1) Penerima KIP Kuliah adalah siswa SMA atau sederajat yang lulus atau akan lulus pada tahun berjalan atau telah dinyatakan lulus maksimal 2 tahun sebelumnya, serta memiliki NISN, NPSN dan NIK yang valid, 2) Memiliki potensi akademik baik, namun memiliki keterbatasan ekonomi yang didukung bukti dokumen yang sah, 3) Siswa SMA/ SMK/ MA atau sederajat yang lulus pada tahun berjalan dengan potensi akademik baik dan mempunyai Kartu KIP atau memiliki Kartu Keluarga Sejahtera atau terdata di Data Terpadu Kesejahteraan Sosial (DTKS) Kemensos, 4) Lulus seleksi penerimaan mahasiswa baru, dan diterima di PTN atau PTS pada Prodi dengan Akreditasi A atau B, dan dimungkinkan dengan pertimbangan tertentu pada Prodi dengan Akreditasi C. Berdasarkan observasi dan wawancara yang telah

dilakukan, ditemukan permasalahan tentang pembagian KIP yang tidak merata dan kurang tepat sasaran. Sehingga penulis akan melakukan *clustering* untuk mengelompokkan siswa penerima KIP agar nantinya dapat digunakan sebagai evaluasi, apakah data penerima sudah merata, apakah data sudah tepat sasaran atau tidak.

Pada penelitian yang dilakukan oleh Syahra, Yohanni. Penjurusan siswa Sekolah Menengah Atas merupakan usaha peningkatan kualitas pendidikan yang langsung berkenaan dengan siswa sebagai salah satu bagian dari peningkatan SDM. Penentuan jurusan yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya dan tidak ada acuan yang yang jelas untuk melakukan pengelompokkan data nilai siswa. Untuk dapat menyelesaikan permasalahan yang ada maka dibuatlah sebuah sistem untuk pengelompokkan nilai siswa untuk penentuan jurusan dengan menerapkan salah satu metode Clustering dan algoritma yang digunakan adalah Algoritma K-Means untuk mengidentifikasi objek yang serupa dengan memperhatikan

beberapa kriteria. Dengan demikian hasil dari pengelompokan yang telah dirancang akan membantu pihak SMA Tamora dalam proses Pengelompokan nilai Siswa untuk penentuan Jurusan yang sesuai dengan kriteria, sehingga pengelompokan dapat dilakukan lebih cepat, tepat dan akurat serta terhindar dari kesalahan. Kemudian penelitian yang dilakukan oleh Fitri Yunita. Proses penerimaan mahasiswa baru Universitas Islam Indragiri menghasilkan data mahasiswa yang sangat berlimpah berupa data profil mahasiswa dan data lainnya. Hal tersebut terjadi secara berulang dan menimbulkan penumpukan terhadap data mahasiswa baru, sehingga mempengaruhi pencarian informasi terhadap data tersebut. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data penerimaan mahasiswa baru di Universitas Islam Indragiri dengan memanfaatkan proses data mining dengan menggunakan teknik Clustering. Algoritma yang digunakan untuk pembentukan cluster adalah algoritma K-Means. K-Means merupakan salah satu metode data non-hierarchical clustering yang dapat mengelompokkan data mahasiswa ke dalam beberapa cluster berdasarkan kemiripan dari data tersebut, sehingga data mahasiswa yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam cluster yang lain. Implementasi menggunakan Rapid Miner 5.3 digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan adalah asal sekolah, program studi dan nilai UAN.

Data mining adalah proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan. *Data mining* merupakan suatu langkah dalam *knowlegde*

discovery in databases (KDD). *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. *Clustering* atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar terhadap objek-objek yang berbeda *cluster*. Objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Objek-objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan objek pada *cluster* yang sama dan memaksimalkan ketidaksamaan pada *cluster* yang berbeda. Kesamaan objek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan objek data, sehingga objek-objek data biasanya dipresentasikan sebagai sebuah titik dalam ruang multidemensi.

Penelitian ini bertujuan untuk membantu meratakan pembagian KIP di Universitas Musi Rawas dan melakukan pengelompokan data dengan lebih baik agar efektif dan efisien untuk membantu pemilihan prioritas penerima KIP. Sehingga pada penelitian saat ini berfokus pada pengelompokan data penerima KIP.

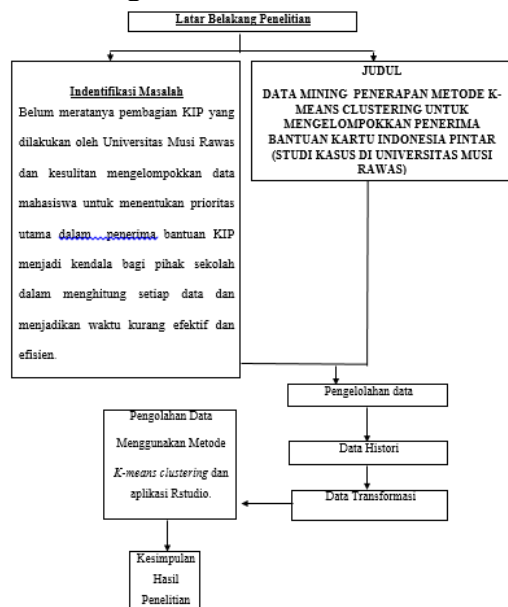
II. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Berdasarkan permasalahan yang akan diteliti, maka penelitian ini menggunakan metode kuantitatif. Metode kuantitatif merupakan metode yang mempunyai struktur tegas dan teratur, maka tahapan dari awal hingga akhir penelitian sudah dapat diramalkan. Penelitian kuantitatif banyak menuntut penggunaan

angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut serta penyajian hasil. Penyajian hasil dalam bentuk gambar, grafik, tabel atau tampilan lain yang *representative* akan meningkatkan serapan pembaca serta mempermudah penyampaian informasi.

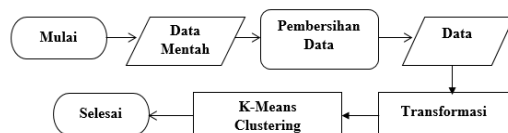
2.2 Kerangka Penelitian



Gambar 2.1 Kerangka Penelitian

2.3 Model Analisa

Pada bagian model analisis berisi alur analisis yang terdiri dari *Processing* data *K-Means Clustering*. Berikut ini Penjelasannya:



Gambar 2.2 Model Analisa

Pada Gambar 2.2 proses *clustering* di mulai dari input data mentah yang berupa data penerima kartu Indonesia pintar yang akan mengalami representasi dari data. Tahap kedua yaitu tahap data pembersihan (*cleaning*) untuk membuang data yang tidak konsisten. Apabila sudah selesai maka didapatkan data yang sudah diproses atau data matang siap untuk ke tahap

selanjutnya. Tahap ketiga yaitu transformasi data, data yang berjenis alfabet di ubah kedalam bentuk angka / numerikal. Tahap ke empat yaitu K-Means Clustering disini tahap K-Means Clustering adalah mengelompokkan data yang sudah ada ke dalam dua kelompok yaitu layak dan tidak layak.

III. HASIL DAN PEMBAHASAN

Hasil yang diperoleh pada penelitian ini adalah *clustering* menggunakan program Rstudio. *Clustering* dilakukan dengan beberapa proses dan *library* dalam program agar algoritma *clustering* dapat dijalankan. Proses *clustering* dilakukan dengan mempersiapkan dataset berupa data mahasiswa penerima KIP Universitas Musi Rawas Tahun 2022. Kemudian mempersiapkan *library* di dalam program Rstudio sesuai dengan keperluan, seperti *cluster*, *factoextra*, *ggplot2*, dan *tidyverse*.

Proses selanjutnya yang dilakukan adalah *import* data dari excel ke Rstudio. *Import* dilakukan dengan mengcopy data dari excel kemudian membentuk kode di dalam Rstudio yang dapat membaca data yang kita copy dengan perintah *read.delim*. Kemudian menampilkan data hasil *import* ke dalam bentuk tabel dan dalam bentuk deskripsi. Setelah data berhasil ditampilkan, langkah selanjutnya adalah melakukan transformasi data. Setelah proses transformasi selesai, langkah selanjutnya yaitu menampilkan data hasil transformasi.

Proses selanjutnya yaitu menentukan jumlah *cluster* dengan *k-means*. Kemudian menampilkan hasil *clustering* dalam bentuk grafik dan tabel. Langkah terakhir yang dilakukan adalah melihat deskripsi hasil data *clustering* berdasarkan *cluster*.

Nilai rata-rata tertinggi untuk program studi berada pada *cluster* 2 sebesar 6,09. Nilai rata-rata tertinggi untuk jumlah saudara berada pada *cluster* 1 sebesar 3,85. Nilai rata-rata tertinggi untuk penghasilan

orang tua berada pada *cluster* 4 sebesar 25000000. Nilai rata-rata tertinggi untuk pekerjaan orang tua berada pada *cluster* 4 sebesar 5. Nilai rata-rata tertinggi untuk status orang tua berada pada *cluster* 3 sebesar 2.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka kesimpulan yang diperoleh yaitu, Pemerataan pembagian KIP untuk Universitas Musi Rawas dapat dilakukan dengan mensortir data mahasiswa berdasarkan program studi, pekerjaan orang tua, penghasilan orang tua, jumlah saudara, dan status orang tua. Kemudian untuk melakukan pembagian KIP dengan efektif dan efisien pihak Universitas Musi Rawas dapat mengelompokkan seluruh mahasiswanya bukan berdasarkan fakultas, namun kelompok ditentukan berdasarkan pekerjaan orang tua, program studi, jumlah saudara, status orang tua dan penghasilan orang tua.

Hasil nilai rata-rata tertinggi untuk program studi berada pada *cluster* 2 sebesar 6,09. Nilai rata-rata tertinggi untuk jumlah saudara berada pada *cluster* 1 sebesar 3,85. Nilai rata-rata tertinggi untuk penghasilan orang tua berada pada *cluster* 4 sebesar 25000000. Nilai rata-rata tertinggi untuk pekerjaan orang tua berada pada *cluster* 4 sebesar 5. Nilai rata-rata tertinggi untuk status orang tua berada pada *cluster* 3 sebesar 2. *Cluster* terbaik dengan ukuran data 20, 22, 3, 1, 22, dan 17 dengan total $\text{between_ss} / \text{total_ss}$ adalah 72,8%. Dan hasil *cluster* terbaik yang diperoleh adalah 6 *cluster*.

V. SARAN

Berdasarkan kesimpulan yang diperoleh, maka penulis dapat memberikan beberapa saran pada penelitian ini, yaitu penelitian yang telah dilakukan diharapkan dapat dijadikan referensi untuk penelitian

selanjutnya dengan tema dan metode yang sama dan Semoga penelitian ini dapat membantu menambah wawasan dan dapat membantu tempat penelitian untuk dapat meningkatkan keefektifan dan efisiensi dalam pengelompokkan mahasiswa penerima KIP.

VI. DAFTAR PUSTAKA

- A. Rahman, S. A. Munandar, A. Fitriani, Y. Karlina, and Yumriani, "Pengertian Pendidikan, Ilmu Pendidikan dan Unsur-Unsur Pendidikan," *Al Urwatul Wutsqa Kaji. Pendidik. Islam*, vol. 2, no. 1, pp. 1–8, 2022.
- N. E. Rohaeni and O. Saryono, "Implementasi Kebijakan Program Indonesia Pintar (PIP) Melalui Kartu Indonesia Pintar (KIP) dalam Upaya Pemerataan Pendidikan," *J. Educ. Manag. Adm. Rev.*, vol. 2, no. 1, pp. 193–204, 2018.
- D. Setya, "Cara Daftar dan Syarat Lengkap KIP Kuliah 2022 untuk PTS," 2022. <https://www.detik.com/edu/beasiswa/d-6217704/cara-daftar-dan-syarat-lengkap-kip-kuliah-2022-untuk-pts-buruan-registrasi-ya#:~:text=Syarat KIP Kuliah 2022,2>.
- L. Vinet and A. Zhedanov, "A 'missing' family of classical orthogonal polynomials," *J. Phys. A Math. Theor.*, vol. 44, no. 8, pp. 1689–1699, 2011, doi: 10.1088/1751-8113/44/8/085201.
- Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- Y. Sarvina, "Pemanfaatan Software Open Source 'R' Untuk Penelitian Agroklimat," *Inform. Pertan.*, vol. 26, no. 1, p. 23, 2017, doi:



- 10.21082/ip.v26n1.2017.p23-30.
- Karsito and W. Monika Sari, "Prediksi Potensi Penjualan Produk Delifrance Dengan Metode Naive Bayes Di Pt. Pangan Lestari," *J. Teknol. Pelita Bangsa*, vol. 9, no. 1, pp. 67–78, 2018.
- F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 91–93, 2013.
- Yuli Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- C. Siadari, "Pengertian Penerapan Menurut Para Ahli," *kumpulanpengertian*, 2020.
<https://www.kumpulanpengertian.com/2020/09/pengertian-penerapan-menurut-para-ahli.html> (accessed Nov. 20, 2022).
- A. Goleman, daniel; Boyatzis, Richard; Mckee and Perdana, "Konsep Character, Capacity, Capital, Condition of Economy dan Colleteral," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- R. Restiyani, "Penerapan Akad Murabah Pada Bank Muamalat Indonesia TBK.," pp. 4–12, 2021.
- Cermati.com, "Kartu Indonesia Pintar: Apa itu, Manfaat, dan Siapa Berhak Menerima," 2015.
<https://www.cermati.com/artikel/kartu-indonesia-pintar-apa-itu-manfaat-dan-siapa-berhak-menerima> (accessed Nov. 20, 2022).
- Y. Syahra, "Penerapan Data Mining Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering," vol. 17, no. 2, pp. 228–233, 2018.
- P. Penerimaan and M. Baru, "Yunita, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri) 1 238," vol. 7, no. September, pp. 238–249, 2018.
- A. Rohmah, F. Sembiring, and A. Erfina, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING ANALYSIS UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS: SMK YASPI M GEGERBITUNG)," pp. 290–298, 2021.
- J. Jendral and A. Yani, "PENERAPAN DATA MINING PADA PENJUALAN MENGGUNAKAN METODE CLUSTERING STUDY KASUS PT . INDOMARCO," pp. 1–11.
- H. O. L. Wijaya, A. A. T. S, A. Armanto, and W. M. Sari, "Prediksi Pola Penjualan Barang pada UMKM XYZ dengan Metode Algoritma Apriori," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 432, 2022, doi: 10.30865/json.v3i4.4200.