



Pengelompokan Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Untuk Meningkatkan Potensi Pemasaran

Daniel Tambun, Sifa Fauziah, Muhtadjuddin Danny

¹Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ¹ Daniel.Tambun21@gmail.com sifa_fauziah@pelitabangsa.ac.id muhtadjuddin.danny@pelitabangsa.ac.id

Email Penulis Korespondensi: daniel.tambun21@gmail.com

Abstrak— Pemanfaatan dataset PMB yang ada melalui pendekatan metode klusterisasi dapat diterapkan dalam menganalisis tingkat penerimaan mahasiswa baru. Model algoritma *K-Means Cluster* yang diterapkan memiliki hasil yang menunjukkan sebuah wawasan baru, yaitu pengelompokan tingkat penerimaan mahasiswa baru berdasarkan 3 kluster, kluster 1 (C0) merupakan tingkat tinggi yang terdiri dari 49 data dari 86 dataset yang diuji dan kluster 2 (C1) adalah merupakan tingkat rendah yang terdiri dari 11 data dari 86 dataset yang diuji dan kluster 3 (C2) adalah merupakan tingkat sedang yang terdiri dari 26 data dari 86 dataset yang diuji. Hasil nilai Davies Bouldin Index atau DBI berdasar pada aplikasi RapidMiner Studio yang didapat dari pengujian data testing, dengan nilai evaluasi Davies -Bouldin Index sebesar 0,769.

Kata Kunci : Data Mining, *K-Means Cluster*, *Klaster*, *PMB*

Abstract – Utilization of the existing PMB dataset through the clustering method approach can be applied in analyzing the rate of acceptance of new students. The K-Means Cluster algorithm model that is applied has results that show a new insight, namely the grouping of new student acceptance rates based on 3 clusters, cluster 1 (C0) is a high level consisting of 49 data from 86 datasets tested and cluster 2 (C1) is a low level consisting of 11 data from 86 datasets tested and cluster 3 (C2) is a medium level consisting of 26 data from 86 datasets tested. The results of the Davies Bouldin Index or DBI value are based on the RapidMiner Studio application obtained from data testing, with a Davies-Bouldin Index evaluation value of 0.769.

Keywords: Data Mining, K-Medoid Cluster, *Klaster*, *PMB*

1. PENDAHULUAN

Kegiatan penerimaan mahasiswa baru merupakan salah satu kegiatan rutin tahunan sebagai media untuk merekrut calon mahasiswa baru. Sekolah tinggi tersebut merupakan salah satu sekolah tinggi di Bekasi yang membuka jalur penerimaan mahasiswa baru melalui ujian mandiri. Pada pelaksanaannya pihak sekolah tinggi menggunakan sistem informasi penerimaan mahasiswa baru sebagai operasional untuk mendata mahasiswa baru yang melakukan pendaftaran dan registrasi ulang.

Penerimaan mahasiswa baru dilaksanakan berdasarkan nilai ujian nasional dan nilai sikap, keterampilan dan sebagian dilaksanakan berdasarkan prestasi non akademik. Hanya saja setelah diterima dengan cara tersebut, mahasiswa mengikuti test tertulis untuk menentukan kelas berdasarkan nilai yang diperoleh ketika mengikuti test tertulis tersebut. Test tertulis yang dilaksanakan meliputi mata pelajaran Bahasa Indonesia, Bahasa Inggris dan Matematika. Ternyata hasil yang didapat menunjukkan bahwa nilai ujian nasional yang diperoleh siswa tidak menjamin bahwa siswa dengan nilai ujian nasional tinggi akan memperoleh nilai yang tinggi pula ketika dilakukan test tertulis, sehingga siswa akan berada pada kelas dengan *grade* yang rendah dan *grade* yang tinggi. Dengan berjalannya proses tersebut, jumlah data mahasiswa baru dari tahun ke tahun mengalami peningkatan, sehingga semakin banyak data yang harus dikelola. Data PMB tersebut meliputi data asal mahasiswa, data asal sekolah, data pekerjaan, prodi yang diambil, serta jalur referensi yang digunakan untuk masuk ke sekolah tinggi.

Saat ini proses pmb di lingkungan sekolah tinggi tersebut telah berjalan dengan menggunakan sistem informasi berbasis web untuk setiap pendaftaran atau registrasi ulang sebagai mahasiswa. Proses penerimaan mahasiswa baru ini merupakan hal yang sangat penting bagi pihak sekolah tinggi terutama manajemen kampus dalam hal ini pihak yayasan. Dengan seiringnya bertambah data penerimaan mahasiswa dari tahun ke tahun terkadang membuat data tersebut semakin banyak dan hanya bersifat sebagai transaksional. Sedangkan, pihak manajemen kampus hanya membutuhkan data berupa laporan yang cepat, mudah, dan tidak mengganggu transaksional dengan tujuan untuk mempermudah pihak manajemen dalam mengambil keputusan yang strategis terkait potensi pemasaran kampus dengan data penerimaan mahasiswa baru.

Maka, dalam hal ini diperlukan sistem yang dapat mendukung keputusan strategis di tingkat sekolah tinggi. Data Mining merupakan elemen penting dalam pengelolaan modal intelektual dan proses pengambilan keputusan untuk membantu para pimpinan dan manajer meningkatkan kinerja perguruan tinggi dalam penerimaan mahasiswa baru. Dalam penelitian yang dilakukan oleh Ika kurniawati menggunakan business intelligence untuk menentukan strategi promosi mahasiswa baru, hasilnya didapatkan strategi promosi bagi calon mahasiswa sesuai dengan program studi yang diminati.

Sedangkan penelitian indri fatma, heru satria tambunan, fitri rizki analisis *cluster* siswa berprestasi dengan

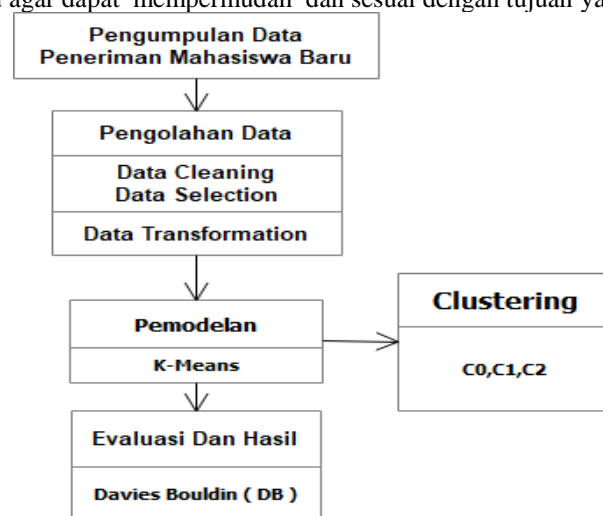
menggunakan data mining yaitu Algoritma *K-Means Cluster*. Yang dimana sebelumnya pihak sekolah masih memakai cara yang manual dalam menentukan siswa yang berprestasi di sekolah tersebut, sehingga membutuhkan waktu yang cukup lama dan hasil yang kurang tepat. *K-Means Cluster* adalah salah satu algoritma yang digunakan untuk klasifikasi atau pengelompokan data, penulis menerapkan algoritma *K-Means Cluster* dalam mengelompokkan siswa yang berprestasi agar mendapatkan hasil yang lebih akurat, cepat dan efektif.

Sedangkan penelitian ramdani udiman, rudianto. Penelitian ini bermaksud untuk menentukan lokasi atau daerah mana saja yang berpotensi untuk mendatangkan mahasiswa baru di perguruan tinggi tersebut. Dengan menerapkan data mining dengan metode clustering mengelompokkan item objek penelitian berdasarkan kemiripan sifatnya, sehingga akan diketahui daerah mana saja yang berpotensi untuk mendatangkan mahasiswa baru. Penentuan lokasi promosi penerimaan mahasiswa baru dengan metode data mining akan memberikan dampak yang bagus dan terarah dalam melakukan promosi, sehingga dapat meningkatkan jumlah mahasiswa baru setiap tahunnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam melakukan analisis dan mencari hasil pada data PMB maka dibuatlah langkah-langkah tersebut untuk meneliti dalam penelitian yaitu agar dapat mempermudah dan sesuai dengan tujuan yang diinginkan



Gambar1 Tahap Penelitian

2.1 Pengujian Metode

Pemodelan data pada penelitian ini dilakukan dengan metode klusterisasi menggunakan algoritma K-Means. Urutan langkah yang dilakukan dalam menggunakan algoritma K-Means adalah sebagai berikut :

- 1 Menentukan jumlah *cluster* yang akan dipakai, dalam penelitian ini dipakai 3 jenis *cluster*, antara lain (C1) untuk data yang memiliki potensi tingkat penjualan yang rendah, (C2) untuk data yang memiliki potensi tingkat yang sedang, dan (C3) untuk data yang memiliki potensi tingkat yang tinggi.
- 2 Menentukan nilai centroid pada tahap awal untuk iterasi ke- 0, dilakukan secara random dengan rumus menentukan target awal k-means untuk mendapatkan target data atau jarak antar kelompok, menggunakan rumus berikut :

Jumlah Data Cn Awal =

Jumlah Class + 1

Keterangan :

$V_{ij} =$

1

$\frac{N_i N_j}{\sum_{k=0}^n X_{kj}}$

$\sum_{k=0}^n X_{kj}$

V_{ij} = Centroid rata-rata cluster ke-i untuk variabel ke-i N_i = jumlah cluster ke -i

i, k = indeks dari cluster

j = indeks dari variabel

X_{kj} = nilai data ke $-k$ variabel ke $-j$ dalam cluster

$$D = \sqrt{(X_i - S_i)^2 + (y_i - t_i)^2}$$

Keterangan :

D = Euclidean Distance e_i = banyaknya objek (x,y) = koordinat objek

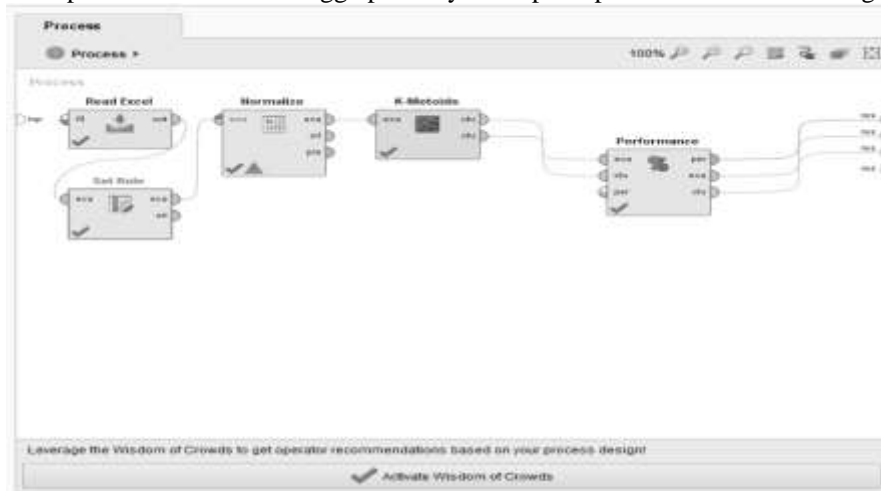
(s,t) = koordinat centroid

3 HASIL DAN PEMBAHASAN

3.1 Pengujian Pada Aplikasi RapidMiner Studio

Pada proses ini metode klusterisasi dengan algoritma K-Means diterapkan untuk pembentukan kelompok cluster dengan keakurasian yang tepat. Dalam penelitian ini penulis menggunakan pengujian perhitungan dengan *tools* Rapid Miner, hasil pengujian yang didapat dengan menggunakan *tools* Rapid Miner adalah dengan tahapan langkah - langkah sebagai berikut :

- 1 Melakukan import data yang diperlukan untuk proses pada tools Rapid Miner. Pada aplikasi Rapid Miner pilih dan klik **Import Data**, kemudian pilih data yang akan dipakai serta kemudian menentukan attribut dan label yang akan digunakan.
- 2 Klik menu **Design**, pada tampilan proses, tambahkan dataset pada folder ke layar tampilan proses. Pada menu **Names & Roles** cari fungsi *Set Role* yang nantinya akan dipakai untuk mengatur role atribut, kemudian drag ke layar tampilan proses.
- 3 Selanjutnya pada menu **Normalization** pilih *Normalize* dan drag ke layar tampilan proses, melalui fungsi ini dapat mengatur normalisasi data yang akan dilakukan dari dataset yang dipakai pada proses ini.
- 4 Kemudian pada menu **Modelling**, dalam submenu **Segmentation**, pilih operator fungsi *k-Means*, untuk menerapkan algoritma **K-Means** terhadap proses klusterisasi yang akan dilakukan. Disini juga ditentukan jumlah cluster yang akan dibentuk dari proses algoritma K-Means tersebut.
- 5 Tambahkan operator fungsi **Performance** untuk melihat hasil evaluasi model dengan kriteria yang ditentukan. Disini penulis menggunakan kriteria Davies Bouldin untuk melihat performa dari model yang dibuat.
- 6 Koneksikan semua perintah tersebut sehingga pada layar tampilan proses terlihat alur sebagai berikut :



Gambar 2. Proses Rapid Miner Klusterisasi K-Means

Setelah dilakukan **Running Process** pada aplikasi RapidMiner Studio, didapatkan hasil pengelompokan cluster data mahasiswa terhadap 50 *record* data yang diproses dan dapat dilihat pada gambar berikut :

Cluster Model

```
Cluster 0: 7 items
Cluster 1: 29 items
Cluster 2: 14 items
Total number of items: 50
```

Gambar 3. Hasil Klaster Data pada Aplikasi RapidMiner Studio



Dapat dilihat melalui hasil cluster model yang terbentuk terhadap 3 cluster yang sudah ditentukan sebelumnya, tiap masing-masing cluster memiliki anggota cluster yang sesuai dengan kelompoknya.

3.2 Pengujian Performa Klastering K-Means

Pengujian performa terhadap model dan algoritma dilakukan dengan maksud mengetahui hasil perhitungan yang dianalisa dan mengukur metode serta algoritma yang digunakan apakah berfungsi dengan baik atau tidak.

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: -0.294  
Avg. within centroid distance_cluster_0: -0.204  
Avg. within centroid distance_cluster_1: -0.235  
Avg. within centroid distance_cluster_2: -0.459  
Davies Bouldin: -0.668
```

Gambar 3. Performa pada Aplikasi RapidMiner Studio

4. KESIMPULAN

Pemanfaatan dataset PMB yang ada melalui pendekatan metode klasterisasi dapat diterapkan dalam menganalisis tingkat penerimaan mahasiswa baru. Model algoritma *K-Means Cluster* yang diterapkan memiliki hasil yang menunjukkan sebuah wawasan baru, yaitu pengelompokan tingkat penerimaan mahasiswa baru berdasarkan 3 klaster, klaster 1 (C0) merupakan tingkat tinggi yang terdiri dari 7 data dari 86 dataset yang diuji dan klaster 2 (C1) adalah merupakan tingkat rendah yang terdiri dari 29 data dari 86 dataset yang diuji dan klaster 14 (C2) adalah merupakan tingkat sedang yang terdiri dari 26 data dari 86 dataset yang diuji. Hasil nilai Davies Bouldin Index atau DBI berdasar pada aplikasi RapidMiner Studio yang didapat dari pengujian data testing, dengan nilai evaluasi Davies - Bouldin Index sebesar 0,769.

REFERENCES

- [1] A. Wibisono, M. Rofik, and E. Purwanto, "Penerapan Analisis Regresi Linier Berganda dalam Penyelesaian Skripsi Mahasiswa," *J. ABDINUS J. Pengabd. Nasant.*, vol. 3, no. 1, p. 30, 2019, doi: 10.29407/ja.v3i1.13512.
- [2] A. S. Nawangsih, Ismasari, "Implementasi Data Mining Untuk Memprediksi Barang Rusak Dengan Menggunakan Algoritma C4.5 Pada Perusahaan Pt Home Center Indonesia," *Progr. Stud. Tek. Inform. Sekol. Tinggi Teknol. Pelita Bangsa*, vol. 9, pp. 34–42, 2019.
- [3] P. S. Ramadhan and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *Sains dan Komput.*, vol. 18, no. 1, pp. 55–61, 2019.
- [4] Suyanto, *Data Mining*. Yogyakarta: Informatika, 2017.
- [5] P. Studi, S. Informasi, J. Sarjana, and A. T. Nopriansyah, "Fakultas ilmu komputer universitas sriwijaya mei 2018," 2018.
- [6] S. Mirghasemi, P. Andrae, M. Zhang, and R. Rayudu, "Severely noisy image segmentation via wavelet shrinkage using PSO and Fuzzy C-Means," *2016 IEEE Symp. Ser. Comput. Intell. SSCI 2016*, 2017, doi: 10.1109/SSCI.2016.7850051.
- [7] Sugiyono, "Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016," *Fik*, vol. 1, no. 1, pp. 1–2, 2016.
- [8] S. Sun, S. Yan, Y. Wang, and Y. Li, "Brain MRI Image Segmentation Based on Improved Fuzzy C-Means Algorithm," *Proc. - 2016 Int. Conf. Smart City Syst. Eng. ICSCSE 2016*, no. 2, pp. 503–505, 2017, doi: 10.1109/ICSCSE.2016.0137.
- [9] N. R. Maulina, I. Surjandari, and A. M. M. Rus, "Data mining approach for customer segmentation in b2b settings using centroid-based clustering," *2019 16th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2019*, no. March, pp. 1–6, 2019, doi: 10.1109/ICSSSM.2019.8887739.
- [10] F. Fengki, "Implementasi Regresi Linear Untuk Memprediksi Lama Waktu Pengiriman Catering Kepada Konsumen Studi Kasus Home Catering Malang," *Cent. Libr. Maulana Malik Ibrahim State Islam. Univ. Malang*, 2018.
- [11] S. Parvathavarthini, N. Karthikeyani Visalakshi, S. Shanthi, and K. Lakshmi, "an Application of Pso-Based Intuitionistic Fuzzy Clustering To Medical Datasets," *ICTACTJ. Soft Comput.*, vol. 8, no. 1, pp. 1531–1538, 2017, doi: 10.21917/ijsc.2017.0213.
- [12] A.-X. Ye and Y.-X. Jin, "A Fuzzy C-Means Clustering Algorithm Based on Improved Quantum Genetic Algorithm," *Int. J. Database Theory Appl.*, vol. 9, no. 1, pp. 227–236, 2016, doi: 10.14257/ijda.2016.9.1.20.
- [13] M. Miftakhul and S. Prihandoko, "Penerapan Algoritma K-Means dan Cure Dalam Menganalisa Pola Perubahan Belanja Dari Retail ke E-Commerce," vol. 7, no. 2, pp. 44–49, 2017.
- [14] O. Villacampa, "(Weka - Thesis) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," *ProQuest Diss. Theses*, no. 63, p. 188, 2015.
- [15] S. Haryati, A. Sudarsono, and E. (2015) Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi



- Mahasiswa Menggunakan Algoritma C4.5,” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [16] A. Khoshkbarchi, A. Kamali, M. Amjadi, and M. Amir Haeri, “A modified hybrid Fuzzy clustering method for big data,” *2016 8th Int. Symp. Telecommun. IST 2016*, pp. 196–201, 2017, doi: 10.1109/ISTEL.2016.7881809.