**IJSA (2025), Vol. 9, No. 1, 117-135**
DOI: https://doi.org/10.29244/ijsa.v9i1p117-135

**Indonesian Journal of Statistics and Its Applications**
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

# K-Prototypes Algorithm for School Indexing in Report Card-Based Student Admissions*

# Algoritma K-Prototypes untuk Indeks Sekolah pada Penerimaan Mahasiswa Baru Jalur Rapor

Ervina Dwi Anggrahini[1], Mohammad Masjkur[1‡], and Utami Dyah Syafitri[1]

[1]Study Program on Statistics and Data Science, IPB University, Indonesia
[‡]corresponding author: masjkur@apps.ipb.ac.id

Copyright © 2025 Ervina Dwi Anggrahini, Mohammad Masjkur, and Utami Dyah Syafitri. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract**

Institut Pertanian Bogor, also known as IPB University, is a state university that was ranked first as the best university in Indonesia by the Ministry of Research and Technology in 2020. It has three main channels in the new student admission selection system. The selection method is called “Seleksi Nasional Berdasarkan Prestasi”. “Seleksi Nasional Berdasarkan Prestasi” is one of the new student admission pathways at IPB University based on report cards without a test. The selection of new student admissions based on report cards requires creating a school index to assess the quality and commitment of each school by grouping schools among “Seleksi Nasional Berdasarkan Prestasi” applicants. One method that can be used is the K-Prototypes algorithm. K-Prototypes can be used to cluster large and mixed-type data (numeric and categorical) by combining distance measures from two non-hierarchical methods, namely the K-Means and K-Modes algorithms. Based on the analysis, the K-Prototypes algorithm yields three optimal clusters, each with distinct characteristics. Cluster 1 is the lowest cluster because it comprises schools with the lowest quality and commitment to new student admissions at IPB University, as indicated by the report card. Cluster 2 has a quality that is not superior to Cluster 3 but is higher than that of Cluster 1. Cluster 3 is the best cluster because it consists of schools that have high quality and commitment to new student admissions at IPB University through the report card route.

**Keywords**: cluster, k-prototypes, new students, school, SNBP

* Received: Sep 2024; Reviewed: Dec 2024; Published: Jun 2025

## 1. Pendahuluan

Institut Pertanian Bogor merupakan salah satu perguruan tinggi negeri di Indonesia yang berdiri sejak tahun 1963. Awal tahun 2019 IPB melakukan *rebranding* dengan meluncurkan perubahan nama Institut Pertanian Bogor (IPB) menjadi IPB University. Meskipun berbasis pertanian, IPB juga dikenal sebagai salah satu perguruan tinggi terbaik di Indonesia dan telah mendapatkan peringkat 1 kampus terbaik se-Indonesia menurut Kemdikbudristek pada tahun 2020. Sebagai salah satu perguruan tinggi negeri terbaik yang berstatus PTN-BH (Perguruan Tinggi Negeri Berbadan Hukum), IPB memiliki tiga jalur seleksi penerimaan mahasiswa baru, yaitu Seleksi Nasional Berdasarkan Prestasi (SNBP), Seleksi Nasional Berdasarkan Tes (SNBT), dan Mandiri (SM-IPB). Di antara jalur tersebut, SNBP merupakan jalur masuk IPB yang menggunakan seleksi nilai rapor. Nilai rapor antar sekolah bervariasi dan tidak bisa dibandingkan secara langsung. Hal ini karena nilai rapor dipengaruhi oleh pola pembelajaran pada masing-masing sekolah. Dalam hal ini perlu dilakukan pemetaaan sekolah-sekolah yang mendaftar pada jalur SNBP berdasarkan beberapa peubah yang melekat dari sekolah. Setelah itu baru dilakukan seleksi pada masing-masing sekolah.

IPB mempunyai histori yang panjang terkait dengan performa lulusan sekolah di tingkat pertama di IPB University. Kualitas internal dari sekolah tersebut juga digunakan dalam pemetaan antara lain rataan skor UTBK serta berapa jumlah siswa tersebut yang diterima UTBK di Perguruan Tinggi lain. Hal tersebut merupakan peubah-peubah yang digunakan untuk pemetaan sekolah.

Pemetaan sekolah dilakukan untuk melihat kualitas dan komitmen sekolah pendaftar SNBP, sehingga sekolah dengan kualitas dan komitmen yang baik akan memiliki peluang lebih besar untuk diterima pada jalur rapor dan sebaliknya. Kurniawan (2016) menyatakan penerimaan calon mahasiswa yang berkualitas dan kompeten diharapkan mampu meningkatkan standar dan kualitas perguruan tinggi itu sendiri.

Seleksi penerimaan mahasiswa baru berdasarkan nilai rapor dilakukan dengan menggerombolkan sekolah-sekolah pendaftar SNBP berdasarkan suatu nilai indeks untuk melihat kualitas dan komitmen masing-masing sekolah. Hal ini dikarenakan setiap sekolah memiliki standar nilai yang berbeda sehingga tidak bisa dibandingkan dan indeks sekolah ini ditentukan berdasarkan indikator keberhasilan belajar di IPB pada semester 1 tahun pertama. Dewantari (2021) dan Aulia (2022) telah melakukan penelitian dengan menggerombolkan sekolah-sekolah pendaftar SNBP pada penerimaan mahasiswa baru tahun 2016−2020. Penelitian tersebut menghasilkan empat gerombol sekolah dengan karakteristik kualitas dan komitmen yang berbeda melalui analisis gerombol dengan algoritma *two-step cluster*. Selain metode *two-step cluster*, metode penggerombolan lain yang dapat digunakan adalah *K-Prototypes*. Algoritma *K-Prototypes* dapat digunakan untuk menggerombolkan data berukuran besar dan bertipe campuran (numerik dan kategorik) dengan menggabungkan ukuran jarak kedua metode non-hierarki, yaitu algoritma *K-Means* dan *K-Modes* (Ganmanah dan Kudus 2021).

Penggerombolan tentang kualitas sekolah menggunakan algoritma *K-Prototypes* berdasarkan data penerimaan mahasiswa baru di IPB University pernah dilakukan oleh Sulastri *et* al. (2021) yang menghasilkan empat gerombol dengan kualitas sekolah terbaik berada pada gerombol empat. Selain itu, Wijayati dan Saputro (2023) telah melakukan penelitian tentang simulasi menggunakan *K-Prototypes* dan menghasilkan kesimpulan bahwa algoritma ini merupakan metode penggerombolan

untuk data bertipe campuran dengan hasil ukuran gerombol terbaik yang dapat dicari menggunakan metode *Elbow*. Fadilah dan Wijayanto (2023) juga melakukan penelitian yang membahas tentang perbandingan metode *One-Hot-Encoding*, *Gower Distance*, dan *K-Prototypes* berdasarkan akurasi pada data penyakit ginjal kronis, diperoleh hasil bahwa *K-Prototypes* mampu menghasilkan *silhouette index* tertinggi dibandingkan dua metode lainnya. Penelitian lain tentang algoritma *K-Prototypes* pernah dilakukan oleh Maulani (2020) yang menggerombolkan desa/kelurahan di Provinsi Banten berdasarkan indikator kemiskinan dan diperoleh gerombol optimal sebanyak 3 dengan rasio keragaman terkecil sebesar 0,114.

Pasaribu (2019) menyatakan meskipun algoritma *K-Prototypes* menghasilkan gerombol tidak unik, tetapi karakteristik gerombol yang dihasilkan relatif sama. Hal ini dikarenakan algoritma *K-Prototypes* mampu menghasilkan nilai rasio paling kecil antara keragaman di dalam gerombol dan keragaman antar gerombol, sehingga algoritma ini dipilih sebagai metode penggerombolan yang akan digunakan dalam penelitian ini.

## 2. Metodologi

### 2.1 Bahan dan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Direktorat Administrasi Pendidikan (DAP) IPB University tahun ajaran 2021–2023. Peubah yang digunakan berupa peubah campuran yang terdiri dari 24 peubah penjelas kontinu, yaitu $X_1$ hingga $X_{24}$, dan 10 peubah kategorik, yaitu $X_{25}$ hingga $X_{34}$. Peubah-peubah yang digunakan pada penelitian ini merujuk pada penelitian yang dilakukan oleh Aulia (2022) dengan rincian peubah dan keterangan peubah kategori perak terdapat pada Tabel 2. Total sekolah yang digunakan dalam analisis ini sebanyak 3981 sekolah.

Tabel 1 Peubah yang digunakan

| Notasi | Nama peubah | Keterangan | Kategori |
|---|---|---|---|
| $X_1 - X_3$ | Persentase daftar ulang SNBP Sarjana tahun 2021−2023 | Numerik | - |
| $X_4 - X_6$ | Persentase daftar ulang SNBP Vokasi tahun 2021−2023 | Numerik | - |
| $X_7 - X_9$ | Persentase daftar ulang SNBT IPB tahun 2021−2023 | Numerik | - |
| $X_{10} - X_{12}$ | Persentase daftar ulang SNBP XY tahun 2021−2023 | Numerik | - |
| $X_{13} - X_{15}$ | Persentase siswa lolos SNBP Sarjana tahun 2021−2023 | Numerik | - |
| $X_{16} - X_{18}$ | Persentase siswa lolos SNBP Vokasi tahun 2021−2023 | Numerik | - |
| $X_{19} - X_{21}$ | Rata-rata skor UTBK sekolah tahun 2021−2023 | Numerik | - |
| $X_{22}$ | Rata-rata nilai rapor IPA sekolah (sarjana) | Numerik | - |
| $X_{23}$ | Rata-rata nilai rapor IPA sekolah (vokasi) | Numerik | - |
| $X_{24}$ | Rata-rata nilai rapor non-IPA sekolah (vokasi) | Numerik | - |

| Notasi | Nama peubah | Keterangan | Kategori |
|---|---|---|---|
| $X_{25} - X_{27}$ | Kategori perak sekolah (sarjana) tahun 2021−2023 | Kategorik | Kategorik PERAK A+ = 11, A=10, A-=9, B+=8, B=7, B-=6, C+= 5, C=4, C-=3, D=2, X=1 |
| $X_{28} - X_{30}$ | Kategori perak sekolah (vokasi) tahun 2021−2023 | Kategorik | Skor: Skor: 11=A+, 10=A, 9=A-, 8=B+, 7=B, 6=B-, 5=C+, 4=C, 3=C-, 2=D, 1=X |
| $X_{31}$ | Akreditasi sekolah | Kategorik | 3 = A, 2 = B, 1 = C, 0 = Lainnya |
| $X_{32}$ | Kepemilikan sekolah | Kategorik | 2=Swasta, 1=Negeri, 0=Lainnya |
| $X_{33}$ | Jenis sekolah | Kategorik | 2 = SMA, 1 = MA, 0 = SMK |
| $X_{34}$ | Daerah asal sekolah | Kategorik | 1=Sumatera, 2=Jawa, 3=Kepulauan Nusa Tenggara, 4=Kalimantan, 5=Sulawesi, 6=Maluku, 7=Papua |

Tabel 2 Keterangan kategori perak IPB

| IP Smt 1 ≥ 2.10 (%) | Kondisi | Kategori |
|---|---|---|
| X ≥ 75 | Ada IP Smt 1 ≥ 3.50 | A+ |
| | 2.75 ≤ IP Smt 1 < 3.50 | A |
| | IP Smt 1 < 2.75 | A- |
| 50 ≤ X < 75 | Ada IP Smt 1 ≥ 3.50 | B+ |
| | 2.75 ≤ IP Smt 1 < 3.50 | B |
| | IP Smt 1 < 2.75 | B- |
| 25 ≤ X < 50 | Ada IP Smt 1 ≥ 3.50 | C+ |
| | 2.75 ≤ IP Smt 1 < 3.50 | C |
| | IP Smt 1 < 2.75 | C- |
| X < 25 | - | D |
| - | - | X |

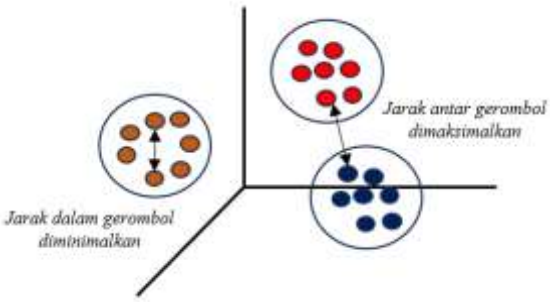
## 2.2 Analisis Gerombol

Analisis gerombol merupakan metode peubah ganda yang bertujuan untuk mengelompokkan *n* objek pengamatan ke dalam *m* gerombol berdasarkan banyaknya *p* peubah, sehingga setiap pengamatan yang terletak dalam satu gerombol memiliki kemiripan sifat yang lebih besar dibandingkan dengan pengamatan yang terletak pada gerombol lain (Johnsons dan Winchern 2002). Analisis gerombol termasuk ke dalam *unsupervised learning* karena algoritma ini bekerja dengan menganalisis data yang tidak berlabel untuk menemukan struktur tersembunyi di dalamnya, data tidak secara eksplisit diberi label ke dalam kelas yang berbeda. Algoritma penggerombolan mengidentifikasi kelompok dalam data dengan mengukur jarak atau kemiripan antar data dan mengelompokkan yang paling mirip ke dalam gerombol yang sama. Analisis gerombol sendiri dibagi menjadi dua metode, yaitu analisis gerombol berhierarki dan analisis gerombol non-hierarki. Analisis gerombol berhierarki adalah metode penggerombolan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan banyaknya gerombol yang diinginkan belum diketahui

jumlahnya (Mattjik dan Sumertajaya 2011). Metode analisis gerombol non-hierarki menurut Johnsons dan Winchern (2002) merupakan metode penggerombolan objek dimana banyaknya gerombol yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur penggerombolan dan metode ini dapat diterapkan pada data dengan ukuran yang lebih besar dibandingkan metode berhierarki.

## 2.3 Ukuran Kemiripan

Ukuran kemiripan merupakan proses pengukuran jarak yang dijadikan acuan dalam menentukan kekuatan hubungan antara dua objek. Seperti yang terlihat pada Gambar 1, semakin meningkat jarak antar dua objek maka semakin berbeda dua objek tersebut, begitu juga sebaliknya semakin kecil jarak antar dua objek maka semakin mirip objek tersebut (Rencher 2002).




Gambar 1 Ilustrasi penggerombolan

Jarak *Euclidean* adalah ukuran jarak yang digunakan untuk tipe data numerik dan biasa digunakan untuk menghitung jarak pada algoritma *K-Means***.** Jarak *Euclidean* antar objek ke-$i$ dan objek ke-$j$ dengan $p$ merupakan banyak peubah numerik adalah sebagai berikut (Salkind 2007):

$$d_{ij} = \left[\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2\right]^{\frac{1}{2}} \quad (1)$$

Keterangan:

$d_{ij}$ = jarak *Euclidean* antara objek ke-$i$ dengan objek ke- $j$

$x_{ik}$ = nilai objek ke-$i$ pada peubah numerik ke-$k$

$x_{jk}$ = nilai objek ke-$j$ pada peubah numerik ke- $k$

$p$ = banyaknya peubah numerik,

Sedangkan ukuran jarak yang digunakan untuk tipe data kategorik adalah jarak *simple matching* yang biasa digunakan pada algoritma *K-Modes* yang ditulis dalam persamaan sebagai berikut (Huang 1998):

$$\boldsymbol{d_{ij} = \sum_{s=1}^{m} \delta(x_{is}, y_{js}),} \quad \boldsymbol{(x_{is}, y_{js})} = \begin{cases} \boldsymbol{0\ (x_{is} = y_{js})} \\ \boldsymbol{1\ (x_{is} \neq y_{js})} \end{cases} \quad (2)$$

Keterangan:

$d_{ij}$ = jarak data kategorik antara objek ke-$i$ dengan objek ke- $j$

$x_{is}$ = nilai objek ke-$i$ pada peubah kategorik ke-$s$

$y_{js}$ = nilai objek ke-$j$ pada peubah kategorik ke-$s$

$m$ = banyaknya peubah kategorik

Berikut contoh perhitungan jarak pada algoritma *K-Modes*:
Misal terdapat 3 objek dengan nilai peubah kategorik latihan, stamina, dan mental seperti yang tercantum pada Tabel 1.

Tabel 3 Data contoh perhitungan jarak data kategorik

| Objek | Latihan | Stamina | Mental |
|---|---|---|---|
| 1 | Rutin | Kuat | Percaya diri (PD) |
| 2 | Jarang | Kuat | Grogi |
| 3 | Jarang | Lemah | Percaya diri (PD) |

Menghitung ketidakmiripan (jarak) antar objek menggunakan jarak *simple matching* sebagai berikut*:*

$d_{ij}$ = $\sum_{s=1}^{m} \delta(x_{is}, y_{js})$

$d_{12}$ = $\delta(x_{11}, y_{21}) + \delta(x_{12}, y_{22}) + \delta(x_{13}, y_{23})$

= $\delta(rutin, jarang) + \delta(kuat, kuat) + \delta(PD, grogi)$

= 1 + 0 + 1 = **2**

$d_{13}$ = $\delta(x_{11}, y_{31}) + \delta(x_{12}, y_{32}) + \delta(x_{13}, y_{33})$

= $\delta(rutin, jarang) + \delta(kuat, lemah) + \delta(PD, PD)$

= 1 + 1 + 0 = **2**

Jadi, jarak data kategorik antara objek ke-1 dengan objek 2 adalah 2 dan jarak objek ke-1 dengan objek 3 juga bernilai 2.

## 2.4 Algoritma *K-Prototypes*

Menurut Pradhana (2022), *K-Prototypes* termasuk dalam metode penggerombolan non-hierarki dan pertama kali dikemukakan oleh Huang pada tahun 1998. Algoritma ini memiliki keunggulan, yaitu mampu mengatasi data bertipe campuran, efektif untuk menangani data yang sangat besar, dan algoritmanya yang tidak terlalu kompleks. Algoritma *K-Prototypes* juga dapat meningkatkan kehomogenan dalam *cluster* dan keheterogenan antar *cluster* dalam mengelompokkan atau menggerombolkan suatu data. Akan tetapi, algoritma *K-Prototypes* juga memiliki kelemahan, yaitu sensitif terhadap penentuan inisialisasi pusat gerombol, sehingga menghasilkan solusi gerombol lokal optimum (Ganmanah dan Kudus 2021), namun kelemahan pada algoritma *K-Prototypes* ini juga dapat diatasi dengan melakukan iterasi untuk mengubah solusi lokal optimum menjadi solusi global optimum (Munthe 2019, 2018).

Ukuran jarak tipe data campuran pada algoritma *K-Prototypes* merupakan penjumlahan jarak *Euclidean* yang dikuadratkan dan ditambahkan dengan hasil perkalian parameter penyeimbang $\gamma$ dengan ukuran jarak tipe data kategorik. Parameter $\gamma$ merupakan penyeimbang perbedaan skala yang digunakan antara peubah numerik dan peubah kategorik. Nilai parameter penyeimbang $\gamma$ yang kecil menunjukkan bahwa penggerombolan didominasi oleh peubah-peubah numerik. Ukuran jarak tipe data campuran yang digunakan pada algoritma *K-Prototypes* dituliskan dalam persamaan berikut:

$$d_{ij} = \sum_{k=1}^{p} (x_{ik} - y_{jk})^2 + \gamma \sum_{s=p+1}^{m} \delta(x_{is}, y_{js}) \qquad (3)$$

Keterangan:

$d_{ij}$ = ukuran jarak tipe data campuran antar objek ke-$i$ dan ke-$j$

$\sum_{k=1}^{p}(x_{ik}-y_{jk})^2$ = ukuran jarak untuk data bertipe numerik
$\sum_{s=p+1}^{m}\delta(x_{is},y_{js})$ = ukuran jarak untuk data bertipe kategorik
$\gamma$ = nilai parameter penyeimbang

Menurut Huang (1998), besarnya nilai parameter $\gamma$ diperoleh dari rata-rata standar deviasi ($\sigma$) seluruh peubah numerik dan berdasarkan simulasi yang telah dilakukan, nilai koefisien gamma ($\gamma$) diperoleh dari $\frac{1}{3}\sigma$ sampai $\frac{p}{3}\sigma$ dengan nilai koefisien $\gamma$ optimal yang biasa digunakan adalah ${}^{p/2}/_{3}\,\sigma$ ($p$ = banyaknya peubah). Oleh karena itu, besarnya nilai parameter penyeimbang gamma ($\gamma$) dipengaruhi oleh banyaknya amatan, peubah kategorik, dan peubah numerik.

## 2.5 Evaluasi Hasil Gerombol

Menurut Han *et al*. (2010), gerombol yang optimal memiliki keragaman antar gerombol yang besar dan keragaman dalam gerombolnya kecil yang berarti bahwa keragaman antar gerombol semakin heterogen dan di dalam gerombol semakin homogen. Evaluasi hasil penggerombolan dapat dihitung berdasarkan rasio antara nilai simpangan baku dalam gerombol ($S_W$) dan nilai simpangan baku antar gerombol ($S_B$). Semakin kecil nilai rasio antara $S_W$ dan $S_B$, semakin baik pula penggerombolan yang dihasilkan. Perhitungan $S_W$ dan $S_B$ untuk peubah bertipe numerik dituliskan dalam persamaan berikut (Bunkers *et al*. 1996):

$$S_W = \frac{1}{C}\sum_{i=1}^{C} S_i \tag{4}$$

Keterangan:
$S_W$ = nilai simpangan baku dalam gerombol
$S_i$ = nilai simpangan baku gerombol ke-$i$
$C$ = banyaknya gerombol yang terbentuk

$$S_B = \left[\frac{1}{C-1}\sum_{i=1}^{C}(\bar{x}_i-\bar{x})^2\right]^{\frac{1}{2}} \tag{5}$$

Keterangan:
$S_B$ = nilai simpangan baku antar gerombol
$\bar{x}_i$ = rata-rata gerombol ke-$i$
$\bar{x}$ = menunjukkan rata-rata keseluruhan gerombol

Terdapat perbedaan untuk perhitungan ukuran keragaman peubah bertipe kategorik, persamaan $S_W$ dan $S_B$ untuk peubah bertipe kategorik dituliskan dalam persamaan berikut (Okada 1999; Kader dan Perry 2007):

$$S_W = [MSW]^{\frac{1}{2}} \tag{6}$$

Nilai $MSW$ dapat diperoleh dari persamaan berikut:

$$MSW = \frac{SSW}{(n-C)} = \frac{1}{(n-C)}\left[\frac{n}{2} - \frac{1}{2}\sum_{i=1}^{C}\frac{1}{n_i}\sum_{j=1}^{m} n_{ji}^2\right] \tag{7}$$

Keterangan:
$S_W$ = nilai simpangan baku dalam gerombol

$MSW$ = rata-rata kuadrat di dalam gerombol
$SSW$ = jumlah kuadrat di dalam gerombol
$n$ = banyaknya seluruh pengamatan
$n_i$ = banyaknya pengamatan pada gerombol ke-$i$
$n_j$ = banyaknya pengamatan pada gerombol ke-$j$
$n_{ji}$ = banyaknya pengamatan kategorik ke-$j$ dan gerombol ke ke-$i$

Banyaknya seluruh amatan $(n)$ dapat dituliskan dengan rumus sebagai berikut:

$$n = \sum_{i=1}^{C} n_i = \sum_{j=1}^{m} n_j = \sum_{i=1}^{C} \sum_{j=1}^{m} n_{ji} \quad (8)$$

Rumus $S_B$ untuk data bertipe kategorik dituliskan dalam persamaan sebagai berikut:

$$S_B = [MSB]^{\frac{1}{2}} \quad (9)$$

Nilai $MSB$ tersebut diperoleh dari persamaan berikut:

$$MSB = \frac{SSB}{C-1} = \frac{1}{C-1}\left[\frac{1}{2}\left(\sum_{i=1}^{C} \frac{1}{n_i} \sum_{j=1}^{m} n_{ji}^2\right) - \frac{1}{2n} \sum_{j=1}^{m} n_j^2\right] \quad (10)$$

Keterangan:
$S_B$ = nilai simpangan baku antar gerombol
$MSB$ = rata-rata kuadrat antar gerombol
$SSB$ = jumlah kuadrat antar gerombol

## 2.6 Prosedur Analisis Data

Pengolahan dan analisis data pada penelitian ini menggunakan software statistik R versi 4.2.2 dan Microsoft Excel. Tahapan analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan praproses data dengan bantuan Microsoft Excel yang meliputi:
   a. Memeriksa kelengkapan data sekolah dalam penerimaan mahasiswa baru IPB jalur SNBP di masing-masing tahun akademik.
   b. Menangani perbedaan skala dengan melakukan standardisasi atau pembakuan data untuk data bertipe numerik di masing-masing tahun akademik.
   c. Melakukan kodefikasi atau pengkodean untuk data bertipe kategorik di masing-masing tahun akademik.
2. Melakukan eksplorasi data untuk memahami struktur dan karakteristik data lebih lanjut. Eksplorasi dilakukan dengan melihat rata-rata persentase siswa mendaftar ulang SNBP dan SNBT di IPB, serta kriteria sekolah berdasarkan kategori perak pada tahun 2018-2023.
3. Melakukan penggerombolan sekolah menggunakan algoritma *K-Prototypes* dengan tahapan sebagai berikut (Gan *et al*. 2007):
   a. Menentukan banyaknya gerombol $(k)$ yang akan dibentuk dengan batas minimun ukuran $k$ adalah dua gerombol dan batas maksimum ukuran $k$ adalah $\sqrt{n}$ , dengan $n$ merupakan banyaknya amatan (Lin *et al*. 2005). Ukuran $k$ yang digunakan dalam penelitian ini, yaitu $k = 2$ hingga $k = 6$.
   b. Menentukan $k$ inisial *prototypes*, yaitu $Z_1$, $Z_2$, …, $Z_k$ sebagai pusat gerombol pada masing-masing gerombol secara acak.

c. Menghitung jarak semua observasi pada set data terhadap inisial gerombol awal, kemudian mengalokasikan semua observasi ke dalam gerombol yang memiliki jarak *prototype* terdekat dengan objek yang diukur.
d. Menghitung titik pusat gerombol baru setelah semua objek dialokasikan. Setelah itu, merealokasikan semua data observasi pada set data terhadap *prototype* yang baru.
e. Apabila titik pusat gerombol tidak berubah atau konvergen, proses algoritma berhenti. Akan tetapi, jika titik pusat masih berubah secara signifikan, proses kembali ke tahap (b) sampai dengan tahap (d) hingga iterasi maksimum tercapai atau tidak ada lagi perpindahan objek.

4. Melakukan evaluasi hasil penggerombolan pada masing-masing $k$ dengan tahapan sebagai berikut:
   a. Menghitung simpangan baku dalam gerombol ($S_W$) dan simpangan baku antar gerombol ($S_B$) pada masing-masing peubah numerik dan peubah kategorik untuk mengevaluasi hasil penggerombolan (Hair *et al*. 2010).
   b. Menghitung nilai rasio $S_W$ total (penjumlahan $S_W$ numerik dan $S_W$ kategorik) dengan $S_B$ total (penjumlahan $S_B$ numerik dan $S_B$ kategorik).
   c. Menentukan gerombol optimal berdasarkan nilai rasio antara $S_W$ dengan $S_B$ terkecil. Gerombol terbaik adalah gerombol dengan nilai rata-rata rasio $S_W$ dan $S_B$ terkecil untuk keseluruhan peubah.

Melakukan visualisasi dan interpretasi hasil sekolah-sekolah pada gerombol optimal berdasarkan karakteristiknya.

## 3. Hasil dan Pembahasan

### 3.1 Deskripsi Data

Data yang bersumber dari Direktorat Administrasi Pendidikan (DAP) IPB menunjukkan terdapat 3.981 sekolah yang telah mendaftarkan lulusannya ke IPB University dari tahun 2021 hingga 2023 melalui jalur rapor, baik pada program pendidikan sarjana maupun diploma. Berdasarkan Tabel 2 dan Tabel 3, mayoritas sekolah yang mendaftar ke IPB University memiliki kategori perak A+, A, dan A-. Kategori perak merupakan pemeringkatan sekolah yang dibuat oleh IPB berdasarkan rekam jejak siswa dan sekolah. Sekolah yang memiliki kategori perak A+ merupakan sekolah yang cenderung terlibat pada semua jalur dan merupakan sekolah yang unggul dalam kurun waktu tertentu (Dewantari 2021), sedangkan kategori perak X diartikan sebagai sekolah yang tidak mendaftarkan lulusannya ke IPB University pada tahun tersebut. Tabel 4 menunjukkan distribusi presentase sekolah yang masuk ke dalam kategori perak sekolah untuk program sarjana dan Tabel 5 menunjukkan hal yang sama untuk program diploma. Sebagian besar sekolah masuk ke dalam kategori perak A+.

Tabel 4 Persentase kategori perak sekolah program pendidikan sarjana

| Tahun | Persentase sekolah per kategori perak | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | |
| 2021 | 59,66 | 30,05 | 3,70 | 2,53 | 1,35 | 0,27 | 0,09 | 0,45 | 0,00 | 1,90 | 100,00 |
| 2022 | 61,35 | 32,62 | 2,57 | 1,24 | 0,53 | 0,18 | 0,00 | 0,00 | 0,00 | 1,51 | 100,00 |
| 2023 | 47,94 | 46,48 | 3,40 | 0,08 | 0,65 | 0,16 | 0,00 | 0,00 | 0,00 | 1,21 | 100,00 |

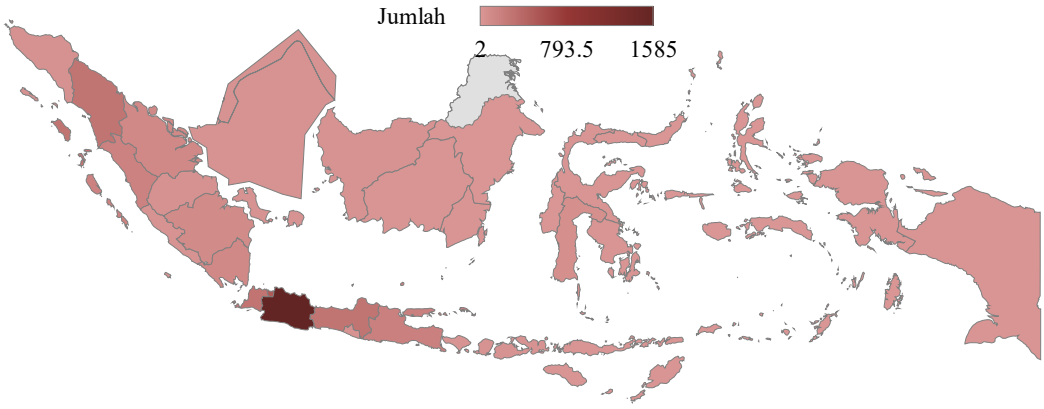
Tabel 5 Persentase kategori perak sekolah program pendidikan diploma

| Tahun | Persentase sekolah per kategori perak | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | |
| 2021 | 79,46 | 1,53 | 3,83 | 3,32 | 4,59 | 6,38 | 0,00 | 0,13 | 0,00 | 0,77 | 100,00 |
| 2022 | 76,90 | 1,44 | 5,51 | 3,02 | 5,64 | 6,56 | 0,00 | 0,13 | 0,00 | 0,79 | 100,00 |
| 2023 | 93,94 | 1,09 | 1,21 | 0,97 | 1,33 | 1,33 | 0,00 | 0,00 | 0,00 | 0,12 | 100,00 |

Selain itu, sekolah yang mendaftar pada jalur rapor sebagian besar merupakan sekolah negeri dengan akreditasi A. Tabel 6 dan Gambar 2 berikut menunjukkan sekolah-sekolah tersebut berasal dari seluruh penjuru Indonesia mulai dari Pulau Sumatera hingga Pulau Papua dengan mayoritas sekolah berasal dari Pulau Jawa.

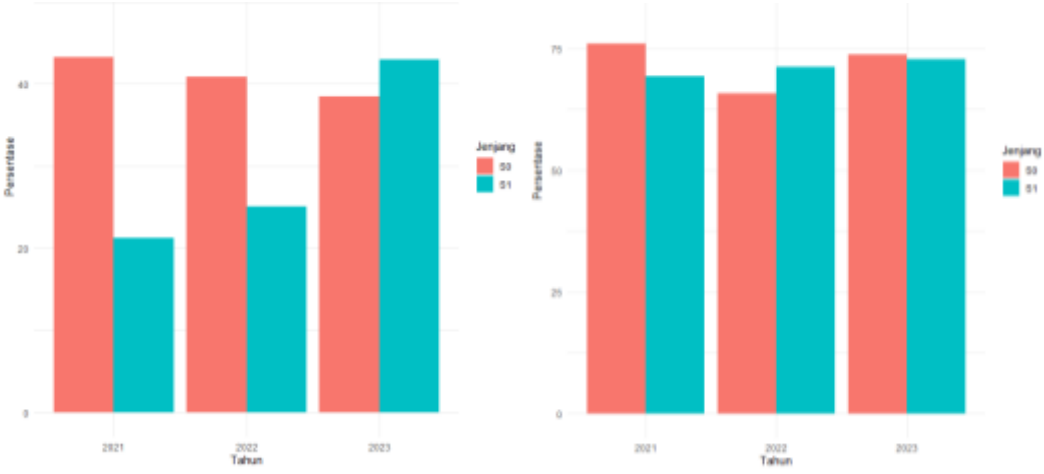
Tabel 6 Sebaran asal sekolah

| Asal | Jumlah sekolah | Persentase |
|---|---|---|
| Pulau Sumatera | 908 | 22,81 |
| Pulau Jawa | 2849 | 71,56 |
| Kepulauan Nusa Tenggara | 54 | 1,36 |
| Pulau Kalimantan | 66 | 1,66 |
| Pulau Sulawesi | 77 | 1,93 |
| Kepulauan Maluku dan Bali | 8 | 0,20 |
| Pulau Papua | 19 | 0,48 |




Gambar 2 Peta tematik sebaran sekolah

Berdasarkan Permendikbudristek No 62 Tahun 2023, bahwa siswa yang diterima di jalur SNBP wajib melakukan registrasi ulang. Pemilihan siswa yang eligible untuk mendaftar SNBP dan pendaftaran SNBP dilakukan oleh sekolah, sehingga komitmen sekolah dapat dilihat berdasarkan siswa yang mendaftar ulang di IPB University. Suatu sekolah dikatakan memiliki komitmen apabila siswa yang lolos pada jalur rapor atau tes di IPB University melakukan daftar ulang. Gambar 3 menunjukkan rata-rata persentase siswa yang melakukan daftar ulang dan lolos melalui jalur rapor pada program pendidikan sarjana (S1) dan program pendidikan diploma (S0). Siswa lolos merupakan persentase rasio jumlah siswa pelamar dengan siswa yang diterima, sedangkan siswa yang melakukan daftar ulang merupakan persentase rasio jumlah siswa lolos dengan siswa yang datang. Rata-rata persentase siswa yang lolos dan mendaftar ulang pada program pendidikan sarjana cenderung mengalami peningkatan setiap tahunnya, sedangkan rata-rata persentase siswa yang lolos pada program pendidikan diploma cenderung mengalami penurunan dan jumlah siswa yang melakukan daftar ulang juga sempat menurun di tahun 2022 hingga akhirnya meningkat kembali di tahun 2023, hal ini dikarenakan pada tahun 2023 jalur masuk pada program pendidikan diploma dilakukan secara nasional, yaitu melalui SNBP.

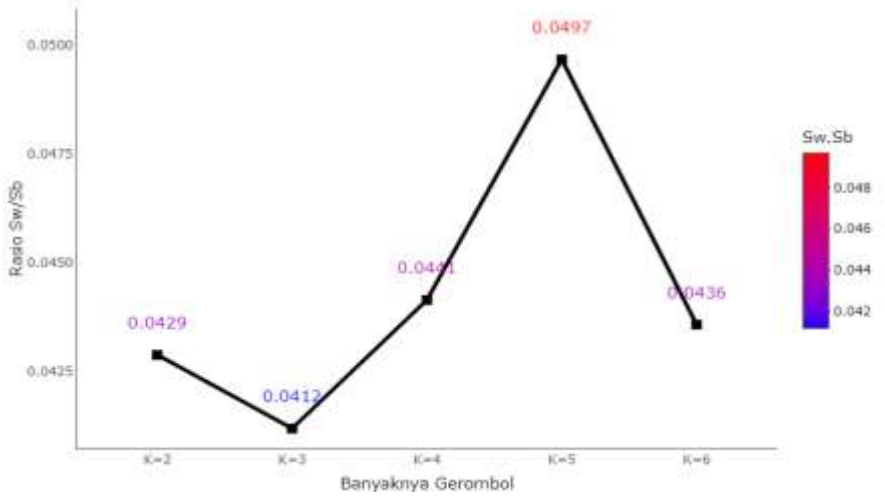



(a) (b)

Gambar 3 Persentase siswa lolos (a) dan daftar ulang (b) pada jalur rapor

### 3.2 Analisis Gerombol dengan Algoritma *K-Prototypes*

Salah satu metode analisis *nonhierarchical clustering* adalah algoritma *K-Prototypes*. Tahapan paling penting dari proses algoritma ini, yaitu menentukan jumlah gerombol awal. Hal tersebut penting karena jumlah gerombol yang berbeda akan menghasilkan hasil gerombol yang berbeda begitupun dengan karakteristiknya, sehingga hasil kesimpulannya juga akan berbeda. Oleh karena itu, perlu dicari jumlah gerombol yang optimal sebelum algoritma *K-Prototypes* dijalankan. Jumlah gerombol optimal ditentukan oleh nilai rasio antara simpangan baku dalam gerombol ($S_W$) dan simpangan baku antar gerombol ($S_B$). Semakin kecil nilai rasionya, semakin baik pula penggerombolan yang dihasilkan, yang berarti keragaman dalam gerombolnya semakin kecil dan kergaman antar gerombolnya semakin besar.




Gambar 4 Rasio $\boldsymbol{S_W}$ dan $\boldsymbol{S_B}$ untuk menentukan jumlah gerombol optimal

Tabel 7 Nilai rasio $S_W$ dan $S_B$

| k | $S_W$ | $S_B$ | $S_W/S_B$ |
|---|---|---|---|
| 2 | 0,7082 | 26,2673 | 0,0429 |
| 3 | 0,6602 | 20,9919 | 0,0412 |
| 4 | 0,6546 | 17,5979 | 0,0441 |
| 5 | 0,6486 | 15,4849 | 0,0497 |
| 6 | 0,6030 | 15,6364 | 0,0436 |

Distribusi penggerombolan pada algoritma *K-Prototypes* akan menghasilkan distribusi yang berbeda setiap setiap kali dilakukan pemrosesan ulang meskipun banyaknya gerombol yang ditentukan sama. Hal ini dapat terjadi karena penentuan inisialisasi pusat gerombol pada setiap pemrosesannya ditentukan secara acak. Oleh karena itu, perlu untuk mengatur dan menentukan nilai *set.seed* pada pemrosesan awal agar distribusi gerombol yang dihasilkan selalu sama untuk setiap proses pengulangan. Pengolahan dalam penelitian ini menggunakan *set.seed*(100) sebagai nilai awal pusat gerombol dengan penentuan jumlah gerombol dimulai dari dua hingga

enam gerombol. Berdasarkan Gambar 4 dapat dilihat bahwa terjadi penurunan nilai rasio $S_W/S_B$ pada saat k = 3, kemudian mengalami kenaikan saat k = 4 dan k = 5, hingga akhirnya menurun secara drastis pada saat k = 6. Nilai terkecil rasio $S_W/S_B$ ditunjukkan ketika k = 3, yaitu sebesar 0,0412. Oleh karena itu, jumlah gerombol optimal yang akan digunakan selanjutnya pada algoritma *K-Prototypes* adalah tiga gerombol. Nilai rasio lengkap $S_W$ dan $S_B$ untuk masing-masing ukuran gerombol dapat dilihat pada Tabel 5. Terbentuknya gerombol pada algoritma *K-Prototypes* juga dipengaruhi oleh parameter penyeimbang gamma ($\gamma$) dan berdasarkan pengolahan menggunakan software R diperoleh nilai parameter penyeimbang $\gamma$ sebesar 2,4443. Nilai ini sama untuk setiap jumlah gerombol dari dua hingga enam dan nilai tersebut diperoleh berdasarkan banyaknya pengamatan, jumlah peubah numerik, dan jumlah peubah kategorik.

Hasil penggerombolan sekolah tahun akademik 2021-2023 ditentukan bahwa gerombol optimal yang dipilih sebanyak tiga gerombol. Berdasarkan Tabel 6, gerombol dua merupakan gerombol dengan anggota sekolah terbanyak, yaitu 2612 sekolah, sedangkan gerombol yang memiliki anggota paling sedikit adalah gerombol satu dengan anggota sebanyak 396 sekolah. Jumlah sekolah yang digunakan dalam penelitian ini terlalu banyak, sehingga semua kode sekolah tiap gerombol tidak dapat ditampilkan dalam diskusi.

Tabel 8 Sebaran anggota gerombol

| Gerombol | Jumlah sekolah dalam gerombol | Persentase (%) |
|---|---|---|
| 1 | 396 | 9,9 |
| 2 | 2612 | 65,6 |
| 3 | 973 | 24,5 |
| Total | 3981 | 100,00 |

## 3.3 Karakteristik Gerombol Berdasarkan Peubah Kategorik

Karakteristik dari masing-masing gerombol dapat dilihat berdasarkan peubah-peubah yang digunakan. Ketiga gerombol yang dihasilkan memiliki karakteristik yang berbeda-beda. Berdasarkan Tabel 7 sampai dengan Tabel 9, dapat dilihat bahwa hampir seluruh sekolah pada gerombol 1 memiliki kategori X yang berarti mayoritas anggota pada gerombol ini merupakan sekolah yang tidak mendaftarkan lulusannya ke IPB University pada tahun tersebut. Adapun kategori A pada gerombol 3 merupakan kategori dengan jumlah sekolah paling banyak di setiap program pendidikan dibandingkan kategori lainnya. Gerombol 2 mayoritas terdiri dari sekolah-sekolah yang memiliki kategori perak X, akan tetapi sekolah dengan kategori perak A pada gerombol ini juga terbilang cukup banyak.

**Tabel 9 Kategori perak sekolah Gerombol 1**

| Tahun | Persentase kategori perak sekolah program pendidikan sarjana | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | |
| 2021 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 100,00 | 100,00 |
| 2022 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 100,00 | 100,00 |
| 2023 | 0,25 | 0,76 | 0,25 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 98,74 | 100,00 |
| Persentase kategori perak sekolah program pendidikan diploma | | | | | | | | | | | | |

| Tahun | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2021 | 0,25 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 99,75 | 100,00 |
| 2022 | 0,25 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 99,75 | 100,00 |
| 2023 | 1,77 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 98,23 | 100,00 |

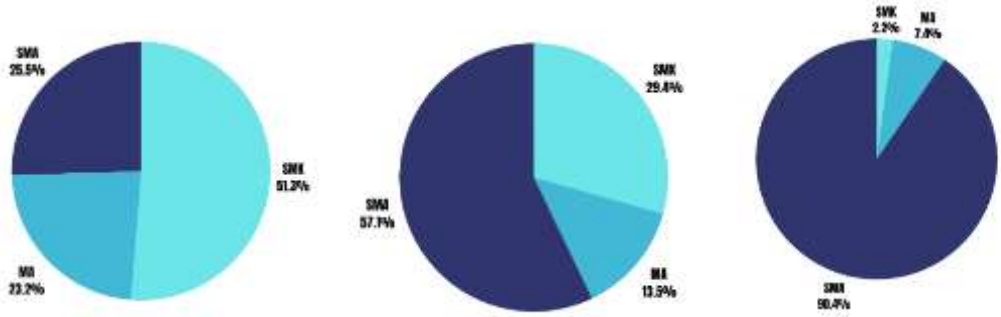
**Tabel 10 Kategori perak sekolah Gerombol 2**

| Tahun | Persentase kategori perak sekolah program pendidikan sarjana | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | |
| 2021 | 57,04 | 20,76 | 1,64 | 2,16 | 1,03 | 0,21 | 0,10 | 0,41 | 0,00 | 0,82 | 15,83 | 100,00 |
| 2022 | 60,43 | 21,79 | 1,64 | 1,03 | 0,41 | 0,21 | 0,00 | 0,00 | 0,00 | 0,72 | 13,77 | 100,00 |
| 2023 | 51,08 | 36,07 | 1,13 | 0,10 | 0,72 | 0,10 | 0,00 | 0,00 | 0,00 | 0,51 | 10,28 | 100,00 |
| Tahun | Persentase kategori perak sekolah program pendidikan diploma | | | | | | | | | | | Total |
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | |
| 2021 | 50,67 | 1,23 | 2,67 | 2,57 | 3,19 | 4,21 | 0,00 | 0,10 | 0,00 | 0,51 | 34,84 | 100,00 |
| 2022 | 44,30 | 1,13 | 4,32 | 2,26 | 4,11 | 4,42 | 0,00 | 0,10 | 0,00 | 0,62 | 38,75 | 100,00 |
| 2023 | 60,02 | 0,92 | 0,92 | 0,72 | 1,13 | 0,92 | 0,00 | 0,00 | 0,00 | 0,10 | 35,25 | 100,00 |

**Tabel 11 Kategori perak sekolah Gerombol 3**

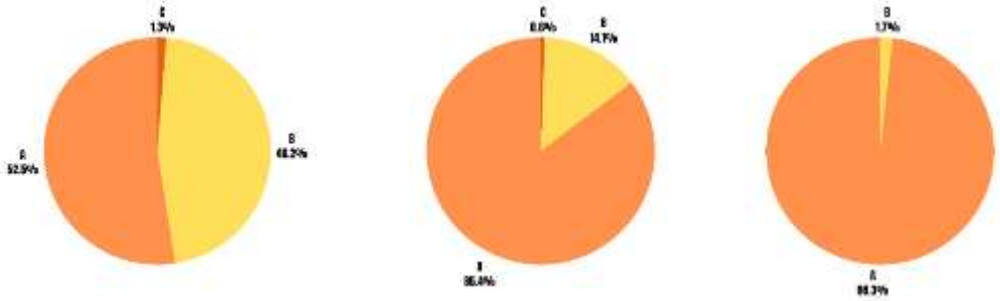
| Tahun | Persentase kategori perak sekolah program pendidikan sarjana | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | |
| 2021 | 4,06 | 5,02 | 0,96 | 0,27 | 0,19 | 0,04 | 0,00 | 0,04 | 0,00 | 0,46 | 88,86 | 100,00 |
| 2022 | 3,98 | 5,97 | 0,50 | 0,15 | 0,08 | 0,00 | 0,00 | 0,00 | 0,00 | 0,38 | 88,82 | 100,00 |
| 2023 | 3,64 | 8,42 | 1,15 | 0,00 | 0,04 | 0,04 | 0,00 | 0,00 | 0,00 | 0,38 | 86,22 | 100,00 |
| Tahun | Persentase kategori perak sekolah program pendidikan diploma | | | | | | | | | | | Total |
| | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | X | |
| 2021 | 4,94 | 0,00 | 0,15 | 0,04 | 0,19 | 0,34 | 0,00 | 0,00 | 0,00 | 0,04 | 94,18 | 100,00 |
| 2022 | 5,90 | 0,00 | 0,00 | 0,04 | 0,11 | 0,27 | 0,00 | 0,00 | 0,00 | 0,00 | 93,57 | 100,00 |
| 2023 | 7,04 | 0,00 | 0,04 | 0,04 | 0,00 | 0,08 | 0,00 | 0,00 | 0,00 | 0,00 | 92,69 | 100,00 |

Selain kategori perak, terdapat peubah kategorik lain yang menjadi karakteristik bagi setiap gerombol, seperti jenis sekolah, akreditasi sekolah, dan status sekolah. Peubah-peubah ini merupakan peubah yang menjelaskan identitas sekolah pada masing-masing gerombol.



(a) (b) (c)

Gambar 5 Sebaran jenis sekolah pada (a) gerombol 1, (b) gerombol 2, (c) gerombol 3

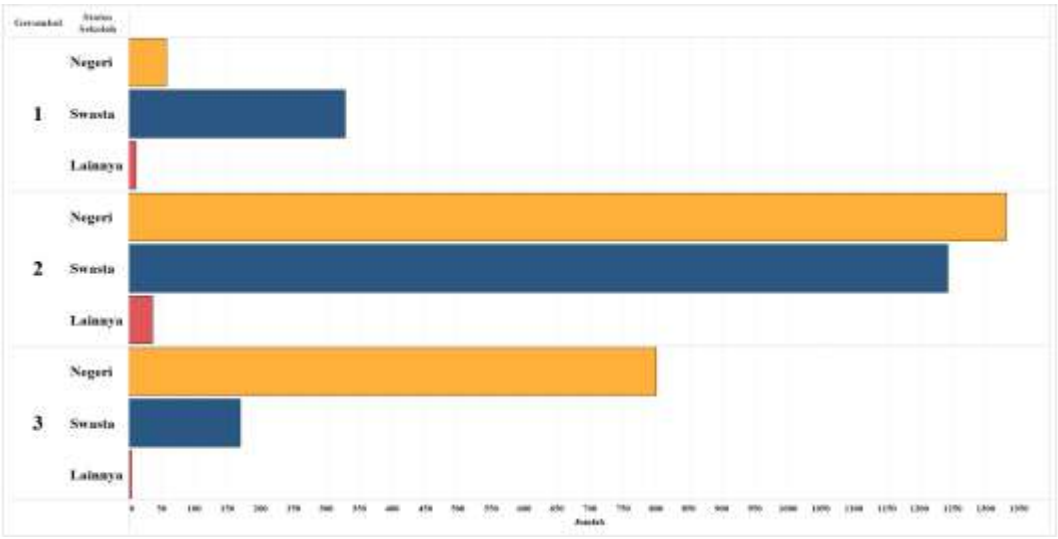


(a) (b) (c)

Gambar 6 Sebaran akreditasi sekolah pada (a) gerombol 1, (b) gerombol 2, (c) gerombol 3

Hasil penggerombolan sekolah berdasarkan peubah jenis dan akreditasi sekolah pada Gambar 5 dan 6 menunjukkan bahwa mayoritas sekolah pada gerombol 1 didominasi oleh sekolah yang berasal dari SMK. Gerombol 2 dan 3 masing-masing didominasi oleh sekolah yang berasal dari SMA. Sekolah yang memiliki partisipasi rendah dalam program SNBP adalah MA, hal ini dapat dilihat dari persentase distribusi sekolah MA yang rendah pada setiap gerombol. Selain itu, sebagian besar sekolah pada gerombol 1, gerombol 2, maupun gerombol 3 yang mendaftar SNBP IPB baik pada program pendidikan sarjana maupun diploma merupakan sekolah dengan akreditasi A.

Sebaran status sekolah pada masing-masing gerombol berdasarkan Gambar 7 menunjukkan bahwa mayoritas sekolah pada gerombol 1 merupakan sekolah berstatus swasta. Adapun sekolah yang berstatus negeri paling banyak berada pada gerombol 3, sedangkan gerombol 2 merupakan satu-satunya gerombol yang sebagian besar terdiri dari sekolah berstatus negeri dan swasta.




Gambar 7 Sebaran status sekolah pada masing-masing gerombol

### 3.4 Karakteristik Gerombol Berdasarkan Peubah Numerik

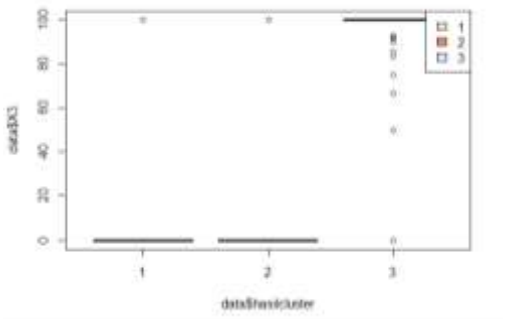
Tabel 12 Statistik deskriptif masing-masing gerombol

| Peubah numerik | Rata-rata ± Simpangan baku | | |
|---|---|---|---|
| | G1 | G2 | G3 |
| X3_Daftar ulang SNBP Sarjana 2023 | 0,25±5,03 | 4,56±20,86 | 80,35±39,49 |
| X6_Daftar ulang SNBP Diploma 2023 | 1,14±10,31 | 5,48±22,15 | 48,47±44,57 |
| X15_Lolos SNBP Sarjana 2023 | 0,25±5,03 | 3,33±16,40 | 45,95±35,48 |
| X18_Lolos SNBP Diploma 2023 | 0,88±8,06 | 3,97±16,31 | 29,00±30,44 |
| X21_Rata-rata skor UTBK 2023 | 236,16±240,71 | 476,58±98,11 | 524,81±30,89 |

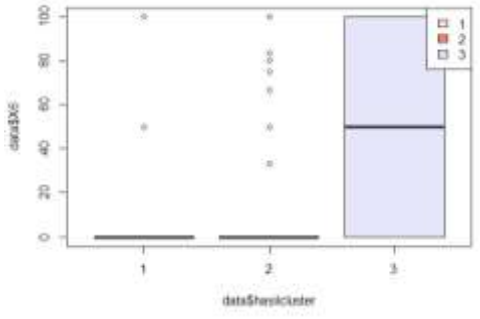
Berdasarkan statistik deskriptif peubah numerik masing-masing gerombol, gerombol 3 merupakan gerombol yang memiliki tingkat lolos dan registrasi tertinggi pada jalur SNBP, baik pada program pendidikan sarjana maupun diploma. Selama program penerimaan mahasiswa baru di IPB University jalur rapor tahun 2023, terdapat 80,35% pandaftar yang melakukan daftar ulang SNBP pada program pendidikan sarjana dan sebanyak 48,47% pandaftar melakukan daftar ulang SNBP pada program pendidikan diploma. Selain itu, rata-rata skor UTBK tertinggi berada pada gerombol 3, dengan kata lain sekolah-sekolah yang masuk dalam gerombol 3 merupakan sekolah yang memiliki kualitas dan predikat sangat baik terkait penerimaan mahasiswa baru di IPB University. Posisi selanjutnya dengan tingkat lolos dan registrasi rendah pada jalur SNBP, namun memiliki rata-rata skor UTBK cukup tinggi adalah gerombol 2. Hal ini berarti gerombol 2 terdiri dari sekolah-sekolah dengan kualitas rendah pada jalur rapor namun memiliki kualitas cukup baik pada jalur tes, sedangkan gerombol 1 merupakan gerombol yang didominasi sekolah-sekolah dengan kualitas kurang baik terkait penerimaan mahasiswa baru di IPB University jalur

rapor, selain memiliki tingkat lolos dan registrasi rendah, gerombol ini juga memiliki rata-rata skor UTBK paling rendah di antara gerombol-gerombol lainnya.

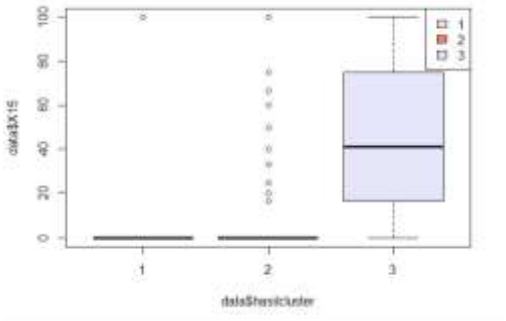
Gambar 8(a) hingga 8(e) di bawah ini juga memperkuat bukti bahwa gerombol 3 merupakan gerombol terbaik karena memiliki sebaran yang lebih besar dibandingkan gerombol lain untuk setiap peubah numeriknya. Berdasarkan gambar tersebut, distribusi data siswa berdasarkan peubah numerik X3, X6, X15, X18, dan X21 memiliki rataan paling besar pada gerombol 3. Gerombol 2 menempati urutan kedua yang memiliki rataan cukup besar dibandingkan dengan gerombol 1, sedangkan gerombol 1 memiliki rataan paling kecil diantara gerombol-gerombol lain. Oleh karena itu sebaran masing-masing peubah yang ditampilkan dalam kotak garis di bawah ini sesuai dengan karakteristik numerik masing-masing gerombol.



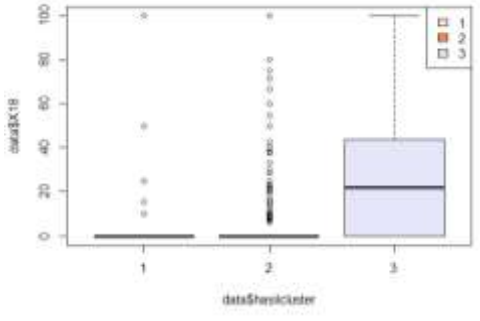
(a) Daftar ulang SNBP Sarjana 2023



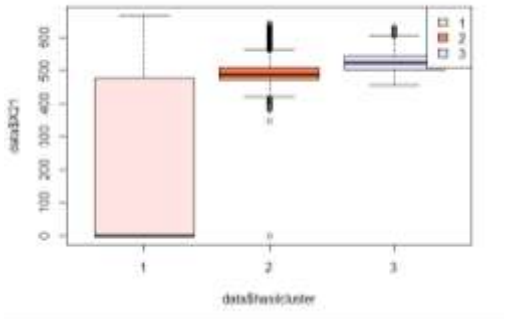
(b) Daftar ulang SNBP Diploma 2023



(c) Lolos SNBP Sarjana 2023



(d) Lolos SNBP Diploma 2023



(e) Rata-rataskor UTBK 2023

Gambar 8 Karakteristik masing-masing gerombol berdasarkan peubah numerik (a) X3, (b) X6, (c) X15, (d) X18, dan (e) X21

## 4. Simpulan dan Saran

Penggerombolan sekolah berdasarkan algoritma K-Protoypes menghasilkan jumlah gerombol optimal sebanyak tiga gerombol. Gerombol pertama terdiri dari 396 sekolah, gerombol 2 terdiri dari 2612 sekolah, dan gerombol 3 terdiri dari 973 sekolah. Berdasarkan analisis peubah numerik dan peubah kategorik, gerombol 3 merupakan gerombol terbaik karena terdiri dari sekolah-sekolah yang memiliki kualitas tinggi pada jalur rapor terkait penerimaan mahasiswa baru di IPB University. Selain itu, pada gerombol ini didominasi sekolah-sekolah berstatus negeri dan berkategori perak A. Gerombol 2 merupakan sekolah-sekolah yang mempunyai portofolio kurang bagus untuk jalur rapot namun rataan skor tes UTBK diatas rata-rata IPB. Gerombol 1 adalah sekolah yang mempunyai track record kurang bagus untuk penerimaan mahasiswa baru di IPB.

Hasil penelitian ini hanya menggerombolkan sekolah-sekolah sebagai alternatif pemetaan indeks mutu tingkat SMA/Sederajat dalam seleksi penerimaan mahasiswa baru jalur rapor di IPB University. Penelitian selanjutnya diharapkan dapat menerapkan metode klasifikasi untuk memperoleh suatu kelompok data yang lebih akurat. Selain itu, analisis korelasi antar peubah dapat dilakukan sebelum analisis lebih lanjut untuk melihat signifikansi masing-masing peubah, misalnya dengan menggunakan metode *Principal Component Analysis* (PCA).

## Daftar Pustaka

Aulia, F. (2022). *Klasifikasi sekolah pada penerimaan mahasiswa baru jalur SNMPTN di IPB University menggunakan metode regresi logistik multinomial* [Skripsi]. Bogor (ID): IPB University.

Bunkers, MJ. Miller, JR. DeGaetano, AT. (1996). Definition of climate region in the northern plains using an objective cluster modification technique. *Journal of Climate*, 9(1): 130-146.

Dewantari, NKM. (2021). *Penggerombolan sekolah pada penerimaan mahasiswa baru jalur SNMPTN di IPB menggunakan metode two-step cluster* [Skripsi]. Bogor (ID): IPB University.

Fadilah, ZR. Wijayanto, AW. Perbandingan metode klasterisasi data bertipe campuran: one-hot-encoding, gower distance, dan *K-Prototypes* berdasarkan akurasi (studi kasus: *chronic kidney disease dataset*). *Journal of Applied Informatics and Computing*, 7(1): 63-73.

Gan, G. MA, C. Wu, J. (2007). *Data Clustering Theory, Algorithms, sand Applications*. Virginia (US): American Statistical Association (ASA).

Ganmanah, M. Kudus, A. (2021). Penerapan algoritma *K-Prototypes* untuk pengelompokkan desa-desa di Provinsi Jawa Barat berdasarkan indikator indeks desa membangun tahun 2020. *Prosiding Statistika*, 7(2): 543-548. doi: http://dx.doi.org/10.29313/.v0i0.28974

Hair, JF. Black, WC. Babin, JB. Anderson, RE. (2010). *Multivariate Data Analysis* .7th ed. New Jersey (US): Pearson Prentice Hall.

Huang, Z. (1997). Clustering large data sets with mixed numerik and categorical values. *Proceeding of the First Pacific Asia Knowladge Discovery and Data Mining Conference*, 21–34.

Huang, Z. (1998). Extension to the K-Means algorithm for clustering large data sets with categorical values. *Data Mining and Knowladge Discovery*, 2(3): 283-304.

Kader, GD. Perry, M. (2007). Variability for categorical variables. *Journal of Statistics Educations*, 15(2): 1–16.

Kurniawan, E. (2016). *Metode topsis untuk menentukan penerimaan mahasiswa baru pendidikan dokter di Universitas Muhammadiyah Purwokerto* [Tesis]. Purwokerto (ID): Universitas Muhammadiyah Purwokerto.

Lin, HJ. Yang, FW. Kao, YT. (2005). An efficient GA-based clustering technique. *Tamkang Journal of Science and Engineering*, 8(2): 113-122.

Maulani. (2020). *Perbandingan desa/kelurahan di Provinsi Banten berdasarkan indikator kemiskinan menggunakan algoritma k-prototypes* [Tesis]. Bogor (ID): IPB University.

Munthe, AD. (2019). *Penggerombolan desa/kelurahan di Provinsi Nusa Tenggara Timur menggunakan algoritma two-step cluster (TSC) dan algoritma k-prototypes* [Tesis]. Bogor (ID): IPB University.

Munthe, AD. Sumertajaya, IM. Syafitri, UD. (2018). Penggerombolan desa/kelurahan berdasarkan indikator kemiskinan dengan menerapkan algoritma TSC dan k-prototypes. *Indonesian Journal of Statistics and Its Applications*, 2(2): 63-76.

Pasaribu, D. (2019). *Segmentasi mahasiswa baru S1 IPB berdasarkan karakteristiknya menggunakan algoritme K-Prototypes* [Skripsi]. Bogor (ID): Institut Pertanian Bogor.

Okada, T. (1999). Sum of squares decomposition for categorical data. *Kwansei Gakuin Studies in Computer Science*, 14: 1-6.

Salkind, NJ. (2007). Encyclopedia of Measurement and Statistics. California (US): SAGE Publications.

Sulastri, S. Usman, L. Syafitri, UD. (2021). *K-Prototypes* algorithm for clustering schools based on the student admission data in IPB University. *Indonesian Journal of Statistics and Its Applications,* 5(2): 228-242. doi: https://doi.org/10.29244/ijsa.v5i2p228-242

Wijayati, R. Saputro, DR. (2023). Clustering data campuran numerik dan kategorik menggunakan algoritme *K-Prototypes*. *PRISMA, Prosiding Seminar Nasional Matematika 2023*, 5(6): 702-706.