

PENERAPAN ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN SISWA LOLOS SNMPTN DI SMAN 8 BANDUNG

Rahmawati¹, Toni Arifin²

¹Universitas Adhirajasa Reswara Sanjaya
email: amarachma97@gmail.com

²Universitas Adhirajasa Reswara Sanjaya
email: toni.tfn19@gmail.com

Abstrak

Pendidikan merupakan hal yang sangat penting, dengan adanya perguruan tinggi maka semakin meningkatkan kualitas pendidikan di Indonesia. Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) merupakan sebuah seleksi nasional yang didasarkan pada nilai rapor dan prestasi. Hal ini juga yang menyebabkan SNMPTN banyak diminati. Siswa yang mengikuti SNMPTN harus menentukan jurusan dan universitas yang tepat agar diterima karena daya tampung SNMPTN terbatas. Pada penelitian ini menggunakan algoritma K-Means dengan tujuan untuk memperoleh informasi mengenai perguruan tinggi mana yang paling banyak diminati oleh siswa siswinya. K-Means merupakan salah satu algoritma yang paling populer karena mudah dan sederhana ketika diimplementasikan. K-Means digunakan untuk mengelompokkan data yang memiliki kemiripan. Data yang digunakan berjumlah 86 data. Berdasarkan validasi menggunakan metrik Davies Bouldin Index maka diperoleh 2 *cluster* yang dinilai dapat mengelompokkan data dengan baik. Performance Vector hasil evaluasi *cluster* yang dibentuk sebanyak 2 *cluster* dengan hasil 0,558. Cluster 1 dengan jumlah 46 items dan cluster 2 dengan jumlah 40 items.

Kata Kunci : K-Means, *Cluster*, SNMPTN

Abstract

Education is very important, with the existence of tertiary institutions, the quality of education in Indonesia is increasingly enhanced. National University Entrance Entrance Selection (SNMPTN) is a national selection based on report card grades and achievements. This also causes SNMPTN to be in great demand. Students who take SNMPTN must determine the right direction and university to be accepted because SNMPTN capacity is limited. In this study using the K-Means algorithm with the aim to obtain information about which universities are most in demand by their students. K-Means is one of the most popular algorithms because it is easy and simple when implemented. K-Means are used to group similar data. The data used amounted to 86 data. Based on validation using the Davies Bouldin Index metric, 2 clusters are obtained which are considered to be able to group data well. Performance Vector results of cluster evaluation formed by 2 clusters with 0.558 results. Cluster 1 with 46 items and Cluster 2 with 40 items.

Keywords: K-Means, *Clusters*, SNMPTN

1. Pendahuluan

Dengan seiring berjalannya zaman, pendidikan merupakan hal yang akan terus berkembang dan sangat penting, di Indonesia sendiri cukup sulit untuk mengimbangi perkembangannya (Ndruru & Limbong, 2018). Fungsi pendidikan yaitu untuk mengembangkan potensi yang dimiliki. Perguruan tinggi merupakan jenjang pendidikan setelah pendidikan menengah atas atau SMA/SMK sederajat. Dengan adanya perguruan tinggi maka semakin meningkatlah kualitas pendidikan di Indonesia (Amri et al., 2019).

Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) adalah sebuah seleksi nasional yang didasarkan pada prestasi akademik seperti nilai rapor dan prestasi lain (Putri et al., 2015). Hal ini menyebabkan SNMPTN menjadi jalur yang banyak diminati, siswa harus menentukan pilihan jurusan dan universitas dengan tepat agar dapat diterima melalui jalur SNMPTN karena dayaampungnya yang terbatas (Prabowo & Sunyoto, 2015).

Sistem seleksi dan rekrutmen mahasiswa baru telah mengalami berbagai perubahan dan penyempurnaan dari waktu ke waktu. Akurasi dan keadilan (*equity*) menjadi acuan dominan dalam seleksi mahasiswa baru. Perubahan ini bertujuan untuk meningkatkan kualitas dan keefektifan sistem pendidikan di Indonesia (Lestari, Anis & Khaeruddin, 2018). Seperti pada data siswa lolos SNMPTN di SMAN 8 Bandung, siswa yang masuk pada seleksi SNMPTN hanya siswa terbaik yang terpilih.

Data mining adalah ilmu yang mempelajari tentang pengumpulan, pembersihan, pengolahan, analisis, dan memperoleh informasi yang berguna dari sebuah data (Saifudin, 2018). Dari sumber lain *data mining* disebut juga sebagai salah satu cara untuk mengubah data menjadi informasi yang berguna serta dapat menghasilkan ilmu baru sebagai wawasan (Ndruru & Limbong, 2018). Salah satu metode yang bisa digunakan yaitu *clustering*, dengan salah satu algoritma yang bisa digunakan adalah K-Means. K-Means adalah salah satu metode pengelompokan data nonhirarki yang mempartisi data. Metode ini mempartisi data dengan karakteristik yang sama ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan

kedalam kelompok data yang lain (Sadewo et al., 2017).

Pada penelitian sebelumnya yang dilakukan oleh (Ndruru & Limbong, 2018) dengan memanfaatkan teknik *data mining* khususnya metode *clustering* untuk mengelompokkan data siswa berdasarkan nilai, dapat menghasilkan informasi mengenai siswa yang berminat untuk memilih jurusan yang diinginkan.

Berdasarkan permasalahan yang ada, maka pada kesempatan kali ini akan dilakukan pengelompokan terhadap data siswa lolos SNMPTN di SMAN 8 Bandung. Data tersebut akan dikelompokkan dengan metode *clustering* menggunakan algoritma K-Means dengan hasil yang diharapkan berupa informasi bagi pihak sekolah untuk mengetahui perguruan tinggi negeri mana paling banyak diambil oleh siswa/i nya.

2. Metode Penelitian

Bagian ini menjelaskan tahapan demi tahapan yang penulis lakukan, dalam menyusun laporan ini agar lebih sistematis dan terarah sesuai dengan yang telah direncanakan.

2.1. Tahapan Penelitian

1. Pengumpulan data
Pada tahap ini akan dijelaskan data yang didapat berasal darimana dan mampu mendeskripsikan data mana saja yang akan digunakan pada penelitian ini.
2. Pengolahan data awal
Pada tahap ini akan dijelaskan mengenai data yang telah layak dilakukan proses pengolahan. Pengolahan data juga berfungsi untuk mempermudah pembentukan model.
3. Model yang diusulkan
Setelah melakukan pengolahan data awal, pada tahap ini akan dijelaskan mengenai model yang diusulkan yang akan digunakan terhadap data yang sudah ada.
4. Eksperimen dan pengujian model
Pada tahap ini akan menjelaskan tentang eksperimen yang dilakukan hingga terbentuknya model, serta akan menjelaskan bagaimana cara menguji model yang terbentuk tersebut.
5. Evaluasi dan validasi hasil
Pada tahap terakhir ini akan dilakukan evaluasi terhadap eksperimen dan pengujian model yang dilakukan sehingga akan mengetahui validasi hasil dari model yang diusulkan.

2.2. Metode Pengumpulan Data

Data yang digunakan merupakan data siswa lolos SNMPTN di SMAN 8 Bandung yang penulis dapatkan dari web sekolah SMAN 8 Bandung yaitu sman8bdg.sch.id. Data dalam penelitian ini sebanyak 86 data yang terdiri dari atribut nama lengkap, kelas, rata-rata, PTN (Perguruan Tinggi Negeri) dan program studi/jurusan.

2.3. Pengolahan Data

Pada tahap ini merupakan tahap untuk memastikan data telah layak dan dapat dilakukan proses pengolahan. Data awal yang penulis peroleh berjumlah 91 data siswa lolos SNMPTN di SMAN 8 Bandung dengan jumlah atribut sebanyak 5 atribut.

Tabel 1. Atribut Keseluruhan

No	Atribut	Keterangan
1	Nama Lengkap	Siswa SMAN 8 Bandung yang lolos mengikuti SNMPTN
2	Kelas	Kelas siswa SMAN 8 Bandung yang lolos mengikuti SNMPTN
3	Rata-rata	Nilai siswa SMAN 8 Bandung yang lolos mengikuti SNMPTN
4	PTN	Perguruan Tinggi Negeri yang dipilih
5	Program Studi/Jurusan	Program studi yang dipilih

Dalam pengolahan data ini dilakukan *data cleaning* atau pembersihan data yang akan digunakan dalam proses perhitungan, data yang tidak lengkap maka akan dihilangkan. Selain itu dilakukan juga *data transformation*, data yang jenisnya nominal seperti kelas, PTN (Perguruan Tinggi Negeri) dan program studi/jurusan harus dilakukan proses inialisasi data terlebih dahulu kedalam bentuk angka atau *numerical*. Dalam melakukan inialisasi dapat dilakukan dengan cara pengurutan angka berdasarkan frekuensinya.

Tabel 2. Tabel Inialisasi Kelas

TABEL KELAS			
No	Kelas	Frekuensi	Inialisasi
1	XII IPS 1	15	1
2	XII IPS 2	10	2
3	XII IPS 3	9	3
...

10	XII MIPA 7	6	10
11	XII MIPA 8	8	11
12	XII MIPA 9	6	12

Berikut merupakan tabel inialisasi dari PTN, yaitu perguruan tinggi negeri yang dipilih oleh siswa SMAN 8 Bandung.

Tabel 3. Tabel Inialisasi PTN

TABEL UNIVERSITAS			
No	Universitas	Frekuensi	Inialisasi
1	UPI	11	1
2	UNPAD	47	2
...
6	ISI	1	6
7	ITB	23	7

Program studi/jurusan merupakan program studi yang dipilih oleh siswa SMAN 8 Bandung. Berikut merupakan tabel inialisasi atribut program studi/jurusan.

Tabel 4. Tabel Inialisasi Program Studi/Jurusan

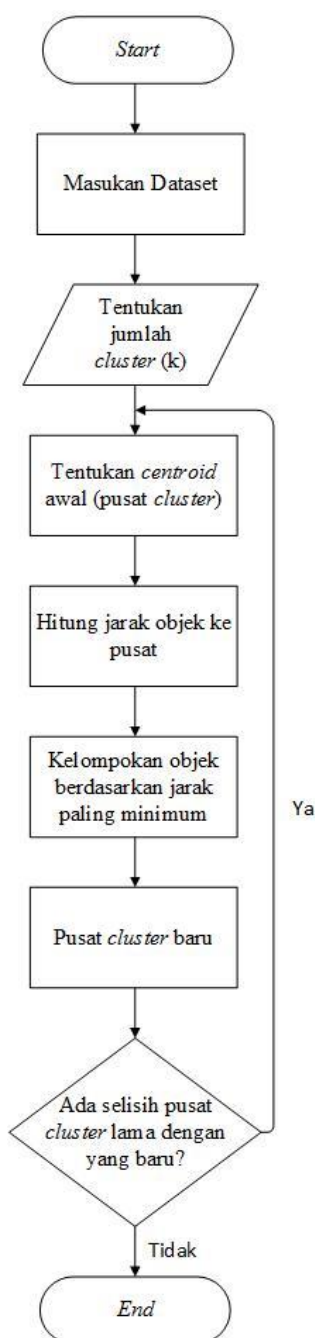
TABEL PROGRAM STUDI/JURUSAN			
No	Program Studi/Jurusan	Frekuensi	Inialisasi
1	Pendidikan Khusus	1	1
2	Akuntansi	2	2
3	Ekonomi Islam	1	3
...
56	Ilmu Aktuaria	1	56
57	FTSL	2	57
58	Biologi	1	58

3. Hasil dan Pembahasan

Pada bagian ini akan dijelaskan pembahasan mengenai hasil penelitian yang penulis lakukan.

3.1. Perhitungan Algoritma K-Means

Pada proses perhitungan algoritma K-Means dilakukan dengan beberapa tahapan :



Gambar 1. Flowchart K-Means

Tahapan pertama dalam perhitungan algoritma K-means adalah menentukan jumlah k (*cluster*), pada penelitian ini dibentuk 2 *cluster* yaitu kelompok *cluster* tertinggi dan rendah. Kemudian menentukan *centroid* awal, *centroid* awal dipilih secara random sesuai dengan atribut Nama Lengkap, Kelas, Rata-rata, PTN dan Program Studi/Jurusan.

Berikut merupakan *dataset* siswa lolos SNMPTN di SMAN 8 Bandung yang telah melewati tahap *preprocessing*

diantaranya tahap *cleaning* atau pembersihan data yang kemudian data yang bertipe nominal dilakukan tahap inisialisasi. Tahap terakhir yaitu transformasi semua data yang telah diinisialisasi.

Tabel 5. Dataset yang sudah di Trasformasi

No	Nama Lengkap	Kelas	Rata-rata	PTN	Program Studi/Jurusan
1	Hesti Ratu Fadilah	1	85.63	1	1
2	Depinkan Zalfaa Faliha	1	88.03	2	2
3	Manda Amelia Saraswati	1	86.9	2	3
4	Adri Mochamad Ihsan	1	86.62	1	4
5	Azizah Muthia Andriana	1	87.25	2	5
6	Ferdiansyah	1	87	2	6
7	Ilham ardian dahlarizandy	1	86.1	2	7
8	Deswita Dwi Lestari	1	86.07	1	8
9	Haidar Husein	1	85.9	1	9
10	Fatimah Az-zahra	1	86.43	1	10
...
77	Alfadli Maulana Siddik	11	86.8	2	54
78	Raden Sylva Adliani	11	88.2	2	55
79	Wafi Fahruzzaman	11	86.8	2	54
80	Sarah Gustiarini Rifdah	11	89.5	7	30
81	Farah Azhariani	12	88.9	7	31
82	M Adhitya Wicaksana	12	86.5	2	56
83	Meisya Paramita Nurhaliza	12	88.69	7	57
84	Canaya Vania Rehansha	12	86.534	2	58
85	Nadira Aby Xavier	12	88.33	7	25
86	Farah Hapsari Yulistiyani	12	89.07	7	57

Tahap selanjutnya yaitu menentukan titik pusat awal dari *cluster*

(centroid) yang dipilih secara random. Pada penelitian ini diambil dari data ke 62 dan 23.

Tabel 6. Centroid Awal

Centroid	No	Nama Lengkap	Kelas	Rata-rata	PTN	Program Studi/Jurusan
Cluster 0	62	Ayumi Yuliani Fuji	8	88,16	2	45
Cluster 1	23	Tesya Salwanda	2	87,33	2	18

Setelah menentukan *centroid* awal, maka tahap selanjutnya yaitu menghitung jarak setiap data yang ada ke *centroid* terdekat untuk menentukan *cluster* menggunakan rumus *Euclidean distance*. Berikut contoh perhitungan jarak ke setiap *centroid*.

➤ Data ke-1 cluster ke-0

$$d(xic0) = \sqrt{a^2 + b^2}$$

$$= \sqrt{(1-8)^2 + (85,63 - 88,16)^2 + (1-2)^2 + (1-45)^2}$$

$$= \sqrt{49 + 6,4009 + 1 + 1936} = \sqrt{1992,4009} = 44,64$$

Dari perhitungan data ke-1 dengan *cluster* ke-0 maka dihasilkan jarak 44,64

➤ Data ke-1 cluster ke-1

$$d(xic0) = \sqrt{a^2 + b^2} =$$

$$\sqrt{(1-2)^2 + (85,63 - 87,33)^2 + (1-2)^2 + (1-18)^2}$$

$$= \sqrt{1 + 2,89 + 1 + 289} = \sqrt{293,89} = 17,14$$

Dari perhitungan data ke-1 dengan *cluster* ke-1 maka dihasilkan jarak 17,14 Hasil perhitungan tersebut yaitu jarak data ke-1 dengan *cluster* ke-0 sebesar 44,64 jarak dan data ke-1 dengan *cluster* ke-1 sebesar 17,14. Berikut Tabel IV.3 yang merupakan perhitungan lengkap dari iterasi ke-1.

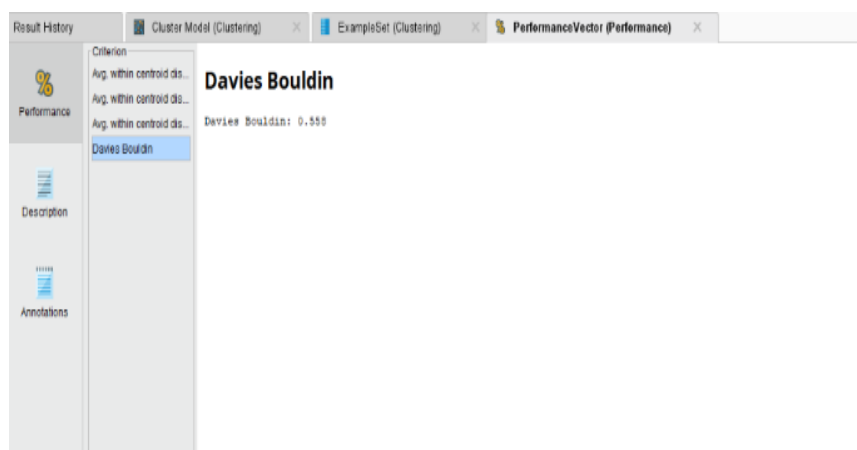
Tabel 7. Hasil Perhitungan dengan Menggunakan Rumus *Euclidean* pada Iterasi ke-1

Data ke-i	Jarak ke Centroid		Jarak Terdekat	Cluster yang diikuti
	C0	C1		
1	44.64	17.14	17.14	C1
2	43.57	16.05	16.05	C1
3	42.60	15.04	15.04	C1
4	41.63	14.09	14.09	C1
5	40.62	13.04	13.04	C1
6	39.64	12.05	12.05	C1
7	38.69	11.11	11.11	C1
8	37.73	10.18	10.18	C1
9	36.76	9.22	9.22	C1
10	35.75	8.17	8.17	C1
...
77	9.58	37.11	9.58	C0
78	10.44	38.09	10.44	C0
79	9.58	37.11	9.58	C0
80	16.15	15.96	15.96	C1
81	15.41	17.22	15.41	C0
82	11.82	39.30	11.82	C0
83	13.61	40.59	13.61	C0
84	13.70	41.24	13.70	C0
85	21.00	13.23	13.23	C1
86	13.63	40.61	13.63	C0

Selanjutnya kelompokan data sesuai dengan jarak *cluster* terdekat. Dari data yang sudah dikelompokan akan didapat *centroid* baru dari hasil rata-rata setiap *cluster*. Lakukan langkah tersebut sampai nilai *centroid* tidak mengalami perubahan.

3.2. Validasi Hasil

Berdasarkan validasi menggunakan metrik *Davies Bouldin Index* pada aplikasi *Rapidminer* maka diperoleh dua *cluster* yang dinilai dapat mengelompokan data dengan baik. *Performance Vector* hasil evaluasi *cluster* yang dibentuk sebanyak dua *cluster* dengan nilai 0,558 yang dapat dilihat pada gambar .



Gambar 2. Performance Vector

Sebelumnya telah dilakukan pengujian dari $k=2$ sampai dengan $k=7$ maka didapat *cluster* yang paling optimal yaitu $k=2$ dengan nilai *Davies Bouldin* sebesar 0,558. *Davies Bouldin Index* dapat dikatakan baik jika nilai semakin kecil atau minimal maka hasil *clustering* yang didapat semakin baik (Kotu & Deshpande, 2015).

Tabel 8. Hasil DBI pengujian $k=2$ sampai $k=7$ pada aplikasi *Rapidminer*

Cluster	Nilai Davies Bouldin
$k=2$	0.558
$k=3$	0.601
$k=4$	0.622
$k=5$	0.692
$k=6$	0.742
$k=7$	0.745

Pada penentuan *cluster* menggunakan metrik *Davies Bouldin Index* digunakan data *centroid* dari iterasi terakhir pada proses *clustering*.

3.3. Pembahasan

Berdasarkan proses *clustering* dengan algoritma K-means menggunakan aplikasi *Rapidminer*, maka menghasilkan informasi bahwa pada *cluster* 0 terdiri dari 46 items dengan Perguruan Tinggi Negeri yang paling diminati adalah UNPAD sebanyak 23 siswa, ITB sebanyak 20 siswa, UNDIP sebanyak 2 siswa dan UPI sebanyak 1 siswa. Pada *cluster* 1 terdiri dari 40 items dengan Perguruan Tinggi Negeri yang

paling diminati adalah UNPAD sebanyak 24 siswa, UPI sebanyak 10 siswa, ITB sebanyak 3 siswa, UGM sebanyak 1 siswa, UNSRI sebanyak 1 siswa dan ISI sebanyak 1 siswa. Berdasarkan uraian tersebut, maka dapat disimpulkan bahwa *cluster* tertinggi berada pada *cluster* 0 yang terdiri dari 46 items dengan dominasi siswa lolos di Perguruan Tinggi Negeri UNPAD dan ITB dan *cluster* 1 menjadi *cluster* rendah yang terdiri dari 28 items dengan dominasi siswa lolos di Perguruan Tinggi Negeri UNPAD dan UPI.

4. Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan data siswa yang lolos SNMPTN di SMAN 8 Bandung menggunakan algoritma K-Means maka diperoleh kesimpulan:

1. Data siswa lolos SNMPTN SMAN 8 Bandung dikelompokkan dengan membuat *design* pemodelan yang dibentuk menjadi 2 *cluster* menggunakan aplikasi *Rapidminer*. Didapatkan hasil pada *cluster* 0 terdiri dari 46 *items* dan *cluster* 1 sebanyak 40 *items*. Maka dapat disimpulkan bahwa *cluster* tertinggi berada pada *cluster* 0 dan *cluster* 1 menjadi *cluster* rendah.
2. Berdasarkan validasi menggunakan metrik *Davies Bouldin Index* pada aplikasi *Rapidminer* maka diperoleh nilai ketepatan sebesar 0,558.

Saran

Guna mewujudkan penelitian selanjutnya yang lebih baik, penulis memberikan beberapa saran, yaitu :

1. Hasil penelitian ini dapat dijadikan evaluasi yang berkaitan dengan minat

- siswa sebelum mendaftar SNMPTN sehingga pihak sekolah dapat mengarahkan siswanya mengambil program studi/jurusan yang diminati
2. Gunakan metrik yang lain agar mendapatkan hasil pengelompokan yang lebih baik
 3. Untuk penelitian selanjutnya gunakan metode lain seperti K-Medoid atau metode lain yang sesuai dengan tipe objek yang digunakan.
 4. Buat aplikasi pencatatan siswa yang lolos SNMPTN sehingga sekolah memiliki data dari setiap tahunnya.

Referensi

- Amri, M. A., Windarto, A. P., Wanto, A., & Damanik, I. S. (2019). Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas Yang Dimiliki Desa. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 674–679. <https://doi.org/10.30865/komik.v3i1.1677>
- Lestari, D., Anis, H., & Khaeruddin, K. (2019). Analisis Hasil Belajar Mahasiswa Jurusan Fisika Berdasarkan Jalur Masuk (Snmptn, Sbmptn Dan Mandiri) Universitas Negeri Makassar. *Jurnal Sains dan Pendidikan Fisika*, 14(3).
- Ndruru, E., & Limbong, R. (2018). Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa'u menggunakan Metode Clustering. *MEANS (Media Informasi Analisa Dan Sistem)*, 3(2), 107–113. http://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/273/pdfdssdx11
- Prabowo, Y. S., & Sunyoto, A. (2015). Sistem Pendukung Keputusan Pemilihan Jurusan SNMPTN Bagi Siswa SMAN 7 Purworejo. *182 CSRID Journal, Vol.6 No.3 Oktober 2014, Hal. 182-195*, 9–10. ejournal.stikom-bali.ac.id/index.php/knsi/article/download/438/91
- Putri, J., Purba, B., Zain, I., Wibowo, W., Ratnasari, V., & Djunaedi, A. (2015). *Pengelompokan Rekam Jejak Sekolah Pada Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) di Institut Teknologi Sepuluh Nopember*. 1(1), 3–7.
- Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 2(1), 60–67. <https://doi.org/10.30743/infotekjar.v2i1.164>
- Saifudin, A. (2018). Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru Di Universitas Pamulang. *10(1)*, 25–36.
- Kotu, V., & Deshpande B. (2015) "Predictive Analytics and Data Mining Concepts and Practice with RapidMiner" Elsevier.