

Analisis Kombinasi Algoritma K-Means Clustering dan TOPSIS Untuk Menentukan Pendekatan Strategi Marketing Berdasarkan Background Target Audiens

Nurus Sarifatul Ngaeni*, Kusri, Kusnawi

Magister Teknik Informatika, Universitas AMIKOM, Yogyakarta, Indonesia

Email: ^{1,*}nurussyarifa@students.amikom.ac.id, ²kusrini@amikom.ac.id, ³khusnawi@amikom.ac.id

Email Penulis Korespondensi: nurussyarifa@students.amikom.ac.id

Submitted: 15/02/2024; Accepted: 23/02/2024; Published: 28/02/2024

Abstrak—Promosi merupakan agenda tahunan STIMIK Tunas Bangsa Banjarnegara. Tujuan dari kegiatan promosi ini yakni untuk menarik mahasiswa baru yang lebih banyak lagi setiap tahunnya. Disisi lain promosi kampus menemui kendala untuk memetakan data pendaftar tahun – tahun sebelumnya sehingga pertimbangan untuk kebijakan promosi baru berdasarkan data sekolah asal alumni atau mahasiswa. Dengan menggunakan algoritma K-Means Clustering, data pendaftar bisa dikelompokkan sesuai dengan latar belakang yang direpresentasikan melalui atribut asal sekolah, pekerjaan orangtua dan daerah asal. Kemudian data diproses menggunakan DSS dengan metode TOPSIS untuk mendapatkan referensi prioritas jenis pemasaran untuk masing – masing cluster. Hasil penghitungan nilai silhouette coefficient untuk lima cluster diperoleh skor 0,426. Sedangkan pada proses perankingan menggunakan metode TOPSIS diperoleh rank pertama terdapat pada cluster 0 dengan skor 0,994110. Tahapan lebih lanjut menggunakan metode Decision Tree untuk mendapatkan output berupa rekomendasi jenis promosi untuk masing – masing cluster. Sebagai contoh, cluster 0 direkomendasikan menggunakan jenis promosi dengan kode P1,P2,P3,P8 dan P9.

Kata Kunci: Sistem Pendukung Keputusan; Pemasaran; Perguruan Tinggi; K-Means Clustering; TOPSIS

Abstract—The promotion is an annual agenda for STIMIK Tunas Bangsa Banjarnegara. The aim of this promotional activity is to attract more new students every year. On the other hand, campus promotion encounters obstacles in mapping applicant data from previous years so that considerations for new promotion policies are based on data from the school of origin of alumni or students. By using the K-Means Clustering algorithm, applicant data can be grouped according to the background represented through the school origin attribute, parents' occupation and place of origin. Then the data is processed using DSS with the TOPSIS method to obtain priority references for marketing types for each cluster. The results of calculating the silhouette coefficient value for the five clusters obtained a score of 0.426. Meanwhile, in the ranking process using the TOPSIS method, the first rank was found in cluster 0 with a score of 0.994110. Further stages use the Decision Tree method to obtain output in the form of recommendations for promotion types for each cluster. For example, cluster 0 is recommended to use promotion types with codes P1, P2, P3, P8 and P9.

Keywords: Decision Support System; Marketing; Perguruan College; K-Means Clustering; TOPSIS

1. PENDAHULUAN

Strategi pemasaran dikatakan cocok mempunyai indikator berupa tercapainya target[1]. Promosi sendiri merupakan bagian dari strategi yang dirancang secara matang oleh perusahaan atau lembaga untuk mencapai tujuan yang disepakati bersama[2]. Melakukan inovasi dalam pemasaran salah satunya bisa dipersiapkan serta dilakukan secara tersistem dengan bantuan teknologi. Tidak terkecuali dalam bidang manajemen pemasaran itu sendiri yang bisa membantu proses analisa berdasarkan data yang dimiliki[3]. Untuk mengetahui kelompok – kelompok data, analisa bisa menggunakan K-Means Clustering[4]. Penggunaan teknologi untuk mencari solusi atau strategi sesuai dengan data yang dimiliki berdasarkan nilai yang paling dekat dengan tujuan serta nilai yang paling jauh bisa menggunakan metode TOPSIS[5].

Penerimaan Mahasiswa Baru (PMB) di kampus STIMIK Tunas Bangsa Banjarnegara merupakan agenda tahunan yang dilaksanakan secara rutin. Tim marketing telah melakukan berbagai jenis pemasaran, termasuk melalui sosial media, pemasangan spanduk, sosialisasi, kerjasama lembaga, beasiswa, dan rekomendasi alumni. Promosi dilakukan hanya berdasarkan program rutin yang sudah dijalankan sebelumnya. Mengingat terbatas sumber daya yang ada, untuk meningkatkan efektivitas dan efisiensi pemasaran, tim memanfaatkan data mining dengan menggunakan metode K-Means Clustering dan metode TOPSIS. Pendekatan ini memungkinkan identifikasi kelompok mahasiswa berdasarkan asal sekolah, pekerjaan orangtua, dan daerah asal. Hasil analisis selanjutnya diproses menggunakan metode TOPSIS untuk mendapatkan urutan cluster terbaik. Untuk mendapatkan keluaran berupa rekomendasi jenis promosi masing – masing cluster, peneliti menggunakan decision tree berdasarkan rumusan yang disetujui oleh Lembaga. Kombinasi metode ini memberikan STIMIK Tunas Bangsa rekomendasi strategi yang sesuai untuk masing – masing cluster, dengan harapan dapat menjadi dasar untuk penentuan kebijakan, termasuk penawaran program dan layanan yang sesuai dengan preferensi target audien. Pendekatan ini juga memungkinkan pengelolaan sumber daya dan anggaran pemasaran dapat difokuskan pada kelompok-kelompok yang memiliki potensi konversi tinggi.

Data mining sudah banyak digunakan sebagai teknik pengolahan data. Dimana data berasal dari pendaftar mahasiswa baru tahun – tahun sebelumnya. Pola atau informasi yang didapatkan dari hasil analisa inilah yang nantinya bisa dijadikan sebagai acuan dalam proses menentukan kebijakan pemasaran. Pengolahan data

mahasiswa baru untuk keperluan pemasaran sudah banyak diamati. Khoironi melakukan penelitian terkait pemanfaatan data mining untuk menyeleksi kandidat penerima beasiswa menggunakan algoritma K-Mean Clustering yang dilanjutkan dengan proses perancangan menggunakan metode AHP[6].

Pengamatan lain dilakukan oleh Oki Oktaviarna Tensao dengan menerapkan data mining dengan algoritma K-Means Clustering untuk membuat kelompok – kelompok mahasiswa baru di STMIK Primakara. Tujuan dari penelitian ini yakni untuk menentukan kebijakan seputar strategi marketing yang paling tepat untuk setiap kelompok mahasiswa. Namun penelitian ini baru berupa pengelompokan, artinya belum bisa memberikan usulan jenis marketing apa yang paling sesuai untuk masing – masing kelompok siswa[7].

Pengamatan lainnya dilakukan oleh Hidayat yang menerapkan algoritma K-Means Clustering untuk mengklasifikasi data jamaah umrah di AET Travel Indonesia. Variabel penelitiannya mencakup usia, jenis kelamin dan daerah asal. Hasil pengelompokan yang didapatkan dalam penelitian ini dijadikan sebagai pengetahuan baru yang mempermudah manager menyusun strategi pemasaran di AET Travel Indonesia. Penelitian mengenai strategi pemasaran di Pendidikan Tinggi dilakukan oleh SUTRISNO & Andriyani yang menggunakan metode pengambilan keputusan berbasis atribut ganda (MADM) dengan metode Simple Additive Weighting (SAW). Tujuan dari penelitian ini yakni berupa rekomendasi asal jurusan apa saja yang paling prospek dijadikan sebagai target promosi[9].

Algoritma K-Means Clustering kemudian dikombinasikan dengan metode SAW untuk mendapatkan rekomendasi lokasi serta jenis promosi yang paling tepat. Penelitian yang dilakukan oleh Kasri dan Jati ini bisa dikembangkan lagi dengan melibatkan beberapa faktor yang mempengaruhi proses pendaftaran mahasiswa baru untuk mendapatkan wawasan yang lebih mendalam terkait optimasi strategi pemasaran. Penelitian selanjutnya dilakukan oleh Andry yang mengembangkan sistem pendukung keputusan untuk menentukan lokasi promosi menggunakan metode Analytical Hierarchy Process (AHP). Tujuan penelitian ini yakni untuk mendapatkan rekomendasi lokasi promosi yang tepat sasaran sehingga manajemen bisa melakukan pemasaran secara lebih efektif dan efisien[11].

Penelitian mengenai strategi promosi juga dilakukan oleh Nanda Ayu Rahmalinda dimana data mahasiswa baru dianalisa menggunakan metode K-Means Clustering untuk menyelaraskan pemilihan strategi promosi[12].

Penelitian lain yang mengkombinasikan dua metode dilakukan oleh Suarnatha yang menggunakan asosiasi algoritma apriori dan perancangan item set terlaris menggunakan TOPSIS. Kedua metode yang digunakan dirasa membuat hasil analisa menjadi lebih bermanfaat[13].

Beberapa kajian yang diuraikan di atas berkaitan dengan tantangan yang dihadapi membuat peneliti menggunakan beberapa metode untuk menentukan jenis pemasaran yang direkomendasikan. Dalam penelitian yang dilakukan oleh Amaliya menyatakan bahwa kombinasi kedua metode antara TOPSIS dan K-Means bisa dibuktikan memberikan hasil yang lebih optimal dalam hal pengelompokan metode pembelajaran berdasarkan bidang keilmuan[14].

Dengan memanfaatkan metode K-Mean Clustering, maka peneliti bisa melakukan identifikasi kelompok – kelompok mahasiswa berdasarkan latar belakang target audiens. Yakni dilihat dari asal sekolah, pekerjaan orangtua, dan daerah asalnya. Selanjutnya hasil analisa akan diolah lagi menggunakan metode TOPSIS untuk membantu proses pemerincian strategi marketing sesuai dengan latar belakang target audien. Untuk mendapatkan output berupa rekomendasi jenis pemasaran untuk masing – masing cluster, digunakan juga penggabungan hasil clustering dan perancangan berdasarkan pohon keputusan yang dibuat. Kombinasi kedua metode ini memungkinkan STMIK Tunas Bangsa mendapatkan rekomendasi strategi yang tepat untuk masing – masing cluster. Latar belakang audiens ini bisa dijadikan sebagai dasar untuk penentuan kebijakan berupa penawaran program serta layanan yang paling sesuai dengan preferensi target audiens.

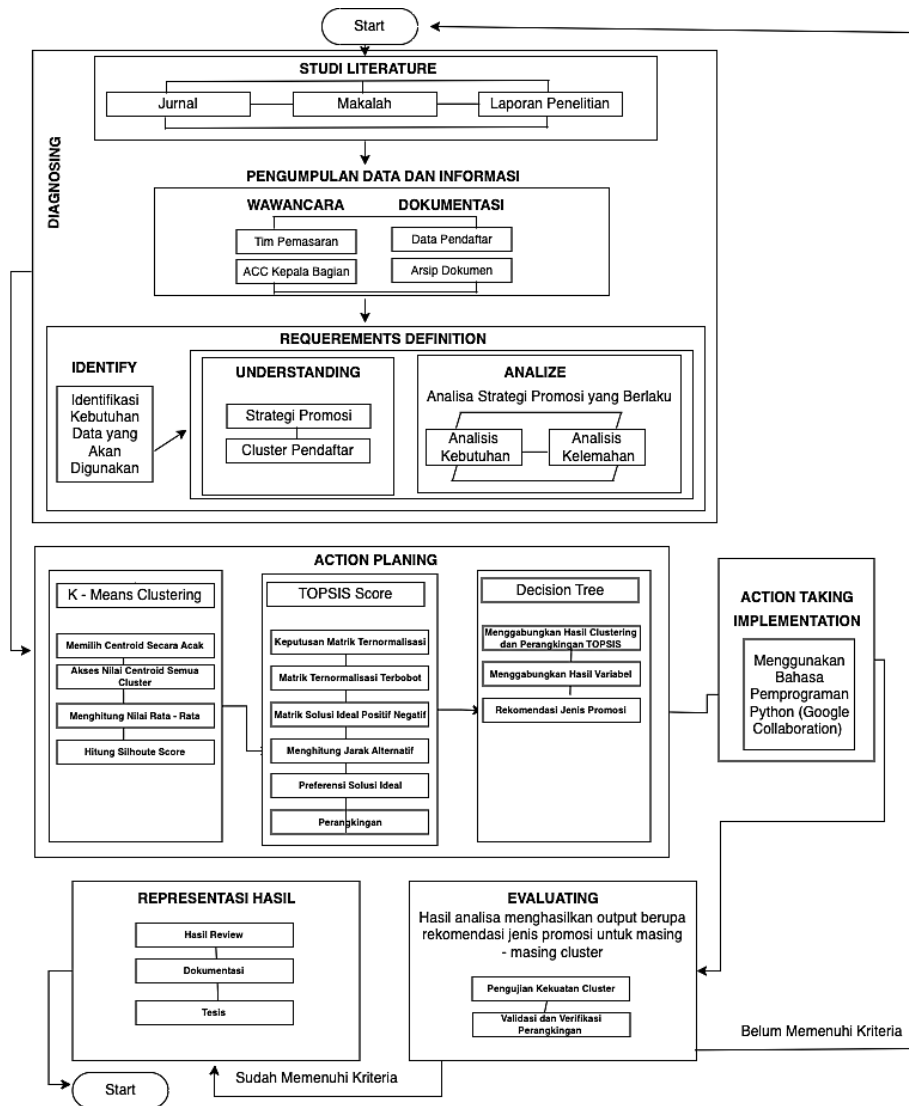
2. METODOLOGI PENELITIAN

2.1 Pengumpulan Data

Dataset penelitian merupakan data pendaftar atau mahasiswa baru yang didapatkan secara studi dokumen melalui bagian pusat administrasi kampus STMIK Tunas Bangsa Banjarnegara. Sedangkan data variabel untuk proses klusterisasi dan perancangan TOPSIS didapatkan dengan metode wawancara melalui divisi pusat marketing kampus STMIK Tunas Bangsa Banjarnegara.

2.2 Tahapan Penelitian

Tahap penelitian diawali dengan *studi literature* dan menentukan kriteria data penelitian. Selanjutnya proses analisis dilakukan menggunakan algoritma K-Means Clustering dan pemerincian menggunakan metode TOPSIS. Sedangkan untuk mendapatkan keluaran berupa rekomendasi jenis pemasaran untuk masing – masing cluster menggunakan pohon keputusan. Tahapan selanjutnya yakni rekomendasi dan kesimpulan seperti yang ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

2.3 Penentuan Masalah dan Kriteria Seleksi

Tujuan yang ingin dicapai dalam penelitian ini yakni mendapatkan informasi mengenai cluster pendaftar serta *output* berupa jenis pemasaran yang paling direkomendasikan untuk masing – masing cluster. Kriteria data pada penelitian ini yakni meliputi variabel asal sekolah sebagai representasi dari latar belakang audiens, pekerjaan orangtua sebagai gambaran kondisi ekonomi keluarga. Sedangkan daerah asal menjadi variabel yang dijadikan sebagai tolak ukur jarak antara sekolah dan tempat tinggal lalu program studi sebagai gambaran pilihan pendaftar. Skenario yang dilakukan yakni peneliti melakukan inisiasi pada setiap variabel data. Nilai inisiasi data untuk setiap subkriteria ditentukan berdasarkan ketentuan dari bagian pemasaran. Misalnya untuk variabel lokasi, peneliti akan menggunakan kecamatan sebagai jarak kilometer dari kampus. Semakin dekat jaraknya maka nilai inisiasinya lebih tinggi. Untuk memudahakan proses penambangan data, kriteria data diinisiasi terlebih dahulu. Inisiasi data untuk masing – masing kriteria terlihat pada Tabel 1.

Tabel 1. Inisiasi Kriteria Data Penelitian

Kriteria	Subkriteria	Nilai
Asal Sekolah (V1)	SMK	8
	SMA	7
	MA/MAN	6
	PKBM	5
Pekerjaan Orangtua (V2)	Pedagang	8
	PNS	7
	Guru Swasta	6
	Karyawan	5
	Petani	4

Kriteria	Subkriteria	Nilai
	Buruh	3
	Yang Lain	2
Program Studi (V3)	S1 Sistem Informasi	1
	S1 Informatika	2
Daerah Asal (V4)	Banjarnegara (1 - 10 km)	8
	Purbalingga (33 km)	7
	Wonosobo (42 km)	6
	Batang (68 km)	5
	Purwokerto (58 km)	4
	Pekalongan (95 km)	3
	Blora (300 km)	2

Kriteria dalam menentukan nilai setiap subkriteria merupakan keputusan kepala bagian beserta tim marketing perguruan tinggi STIMIK Tunas Bangsa Banjarnegara. Inisiasi subkriteria asal sekolah diambil dari data pendaftar terbanyak dan evaluasi tingkat pemahaman mahasiswa mengikuti proses pembelajaran. Inisiasi pekerjaan orangtua menggunakan nilai 2 hingga 8. Dimana tingkat subkriteria paling tinggi diurut dari pekerjaan yang mempunyai kondisi ekonomi lebih baik dan lancar administrasi. Inisiasi variabel program studi terdiri dari dua jurusan yang saat ini tersedia, dimana nilai 2 diberikan untuk program studi dengan jumlah mahasiswa lebih banyak. Data inisiasi asal daerah menggunakan satuan jarak kilometer (km) dimana nilai tertinggi diberikan pada jarak yang paling dekat dengan jangkauan kampus. Pemilihan wilayah ini merupakan daerah asal para pendaftar mahasiswa baru.

Strategi promosi disajikan pada gambar tabel 2, hal ini sesuai dengan teknik promosi yang dilakukan oleh perguruan tinggi.

Tabel 2. Inisiasi Kriteria Data Penelitian

Kode	Jenis Promosi
P1	Instagram
P2	Facebook
P3	Brosur
P4	Spanduk/Flyer
P5	Sosialisasi ke Sekolah
P6	Beasiswa
P7	Kerjasama Lembaga
P8	Seminar
P9	Rekomendasi Teman/Alumni

Rekomendasi pohon keputusan tersaji pada tabel 3 di bawah. Dimana pohon keputusan ini dibuat berdasarkan wawancara dengan ketua departemen dan tim marketing. Pohon keputusan ini sudah menentukan nilai yang nantinya dicapai. Misalkan untuk kriteria V1, V2, V3, dan V4 dengan rentang nilai 2-4, 4, 8, 8, 1 maka untuk jenis promosi yang direkomendasikan yakni P1,P2,P6,P8,P9. Faktor jarak adalah salah satu pertimbangan penting dalam melaksanakan promosi. Dimana semakin jauh jangkauannya maka butuh semakin banyak sumber daya yang dikerahkan. Asal sekolah juga berpengaruh dalam menentukan strategi promosi, dimana latar belakang siswa yang berasal dari SMK/SMA/MA sederajat biasanya mempunyai kemampuan mengikuti proses pembelajaran yang cukup beragam. Lalu selanjutnya yakni faktor pekerjaan orangtua dimana data ini menjadi representasi kondisi ekonomi keluarga mahasiswa. Berdasarkan data yang diperoleh, pekerjaan orangtua paling banyak yakni petani yang mana tergolong penghasilan menengah ke bawah.

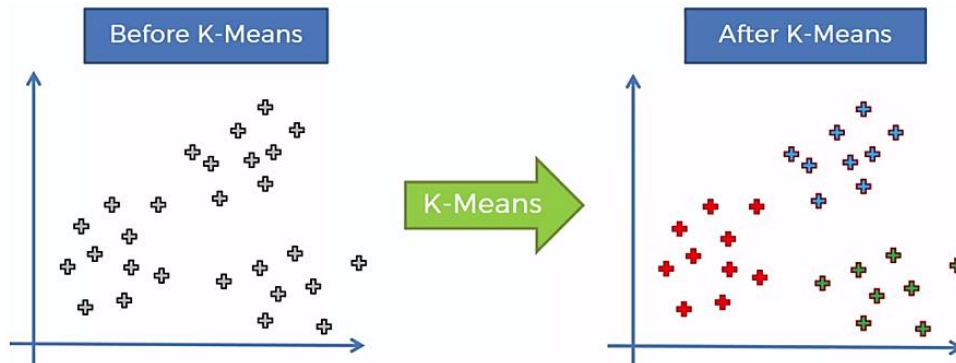
Tabel 3. Pohon Keputusan Penentuan Strategi Promosi

min_daerah_asal	max_daerah_asal	asal_sekolah	pekerjaan_orangtua	program_studi	kode_promosi
2	4	8	8	1	P1,P2,P6,P8,P9
2	4	8	8	2	P1,P2,P6,P8,P9
2	4	8	7	1	P1,P2,P8,P9,P7
2	4	8	7	2	P1,P2,P8,P9,P7
2	4	8	6	1	P1,P2,P6,P8,P9
2	4	8	6	2	P1,P2,P6,P8,P9
2	4	8	5	1	P1,P2,P8,P9
2	4	8	5	2	P1,P2,P8,P9
2	4	8	4	1	P1,P2,P6,P8,P9

min_daerah_asal	max_daerah_asal	asal_sekolah	pekerjaan_orangtua	program_studi	kode_promosi
2	4	8	4	2	P1,P2,P6,P8,P9
2	4	8	3	1	P1,P2,P6,P8,P9
2	4	8	3	2	P1,P2,P6,P8,P9
2	4	8	2	1	P1,P2,P8,P9
2	4	8	2	2	P1,P2,P8,P9
2	4	7	8	1	P1,P2,P6,P8,P9
2	4	7	8	2	P1,P2,P6,P8,P9
2	4	7	7	1	P1,P2,P8,P9,P7
2	4	7	7	2	P1,P2,P8,P9,P7
2	4	7	6	1	P1,P2,P6,P8,P9
2	4	7	6	2	P1,P2,P6,P8,P9
2	4	7	5	1	P1,P2,P8,P9
2	4	7	5	2	P1,P2,P8,P9
2	4	7	4	1	P1,P2,P6,P8,P9
2	4	7	4	2	P1,P2,P6,P8,P9
2	4	7	3	1	P1,P2,P6,P8,P9
2	4	7	3	2	P1,P2,P6,P8,P9
2	4	7	2	1	P1,P2,P8,P9
2	4	7	2	2	P1,P2,P8,P9
2	4	6	8	1	P1,P2,P6,P8,P9
2	4	6	8	2	P1,P2,P6,P8,P9
2	4	6	7	1	P1,P2,P8,P9,P7
2	4	6	7	2	P1,P2,P8,P9,P7
2	4	6	6	1	P1,P2,P6,P8,P9
2	4	6	6	2	P1,P2,P6,P8,P9
2	4	6	5	1	P1,P2,P8,P9
2	4	6	5	2	P1,P2,P8,P9
2	4	6	4	1	P1,P2,P6,P8,P9
2	4	6	4	2	P1,P2,P6,P8,P9
2	4	6	3	1	P1,P2,P6,P8,P9
2	4	6	3	2	P1,P2,P6,P8,P9
2	4	6	2	1	P1,P2,P8,P9
2	4	6	2	2	P1,P2,P8,P9
2	4	5	8	1	P1,P2,P6,P8,P9
2	4	5	8	2	P1,P2,P6,P8,P9
2	4	5	7	1	P1,P2,P8,P9,P7
2	4	5	7	2	P1,P2,P8,P9,P7
2	4	5	6	1	P1,P2,P6,P8,P9
2	4	5	6	2	P1,P2,P6,P8,P9
2	4	5	5	1	P1,P2,P8,P9
2	4	5	5	2	P1,P2,P8,P9
2	4	5	4	1	P1,P2,P6,P8,P9
2	4	5	4	2	P1,P2,P6,P8,P9
2	4	5	3	1	P1,P2,P6,P8,P9
2	4	5	3	2	P1,P2,P6,P8,P9
2	4	5	2	1	P1,P2,P8,P9
2	4	5	2	2	P1,P2,P8,P9

2.4 Pengelompokan K-Means

Algoritma K-Means Clustering ini digunakan untuk membuat cluster - cluster yang mempunyai kemiripan satu dengan lainnya. Cara kerja metode ini ditunjukkan melalui gambar 1 berikut :



Gambar 2. Algoritma K-Means Clustering

Sumber : <https://medium.com/@kurniasp/k-means-clustering-using-scikit-learn-in-python-51073b6f51e5>

K-Means Clustering adalah salah satu jenis metode dalam Data Mining yang digunakan untuk membuat clustering. Dimana objek yang mempunyai kemiripan berdasarkan karakteristik tertentu yang sudah ditetapkan. Hasil dari pengelompokan data ini yakni dijadikan sebagai acuan dalam menentukan strategi tertentu, misalnya saja dalam hal strategi pemasaran. K-Means Clustering merupakan algoritma yang mudah dan efektif dimanfaatkan untuk mengetahui cluster atau kelompok – kelompok data. [15]

K-Means Clustering merupakan salah satu metode clustering yang bersifat non hirarki. Cara kerja yang dilakukan untuk memisahkan antar data yakni dengan memasukkan data satu dengan lainnya yang mempunyai nilai kedekatan. Untuk tahapan melakukan clustering menggunakan algoritma K-Mean Clustering yang dipaparkan dalam penelitian D.Fakhri dkk[16] :

1. Peneliti menentukan jumlah cluster yang nantinya akan dibentuk
2. Pilih centroid awal secara random
3. Menghitung alokasi data berdasarkan jumlah cluster yang dibentuk. Kedekatan antara satu objek dengan objek lainnya akan ditentukan oleh jarak. Artinya mana objek yang paling dekat akan menentukan suatu data masuk ke dalam cluster tertentu.
4. Menghitung nilai pusat cluster (centroid) yang merupakan nilai rata – rata atas semua objek data dalam sebuah cluster.
5. Lakukan kembali alokasi masing – masing objek menggunakan centroid baru sampai nilai pusat cluster tidak berubah.
6. Apabila nilai pusat cluster masih berubah atau ada objek yang berpindah grup maka proses akan dikembalikan pada menghitung jarak antara objek ke centroid.

Perhitungan K-Means dimulai dengan menentukan centroid awal secara acak. Berdasarkan wawancara dengan bagian pemasaran, pemilihan centroid awal yaitu nilai $k=5$. Nilai ini muncul berdasarkan jumlah lokasi promosi yang akan didatangi oleh tim penerimaan mahasiswa baru nantinya.

2.5 Perangkingan TOPSIS

Sebagai metode yang digunakan dalam pengambilan keputusan, TOPSIS ini pertama kali diperkenalkan oleh Yoon dan Hwang pada tahun 1981 untuk menangani pengambilan keputusan bersifat multikriteria. Metode TOPSIS sendiri sudah banyak diaplikasikan untuk keperluan pengambilan keputusan di beragam sektor. Beberapa contohnya yakni diterapkan pada aplikasi yang menunjang keputusan terkait investasi, penilaian performa karyawan, evaluasi pemasaran bahkan terkait proses perancangan robot. Contoh penggunaan metode TOPSIS yakni untuk menentukan karyawan terbaik berdasarkan kriteria tertentu secara objektif[17].

TOPSIS merupakan salah satu metode yang digunakan untuk pengambilan keputusan bersifat multikriteria atau alternatif pilihan yang merupakan alternatif dari jarak paling kecil dari solusi positif dan jarak paling besar dari solusi ideal negatif dilihat dari sudut pandang geometris menggunakan jarak Euclidean. Dan yang menjadi catatan bahwa alternatif dengan jarak paling kecil yang berasal dari solusi ideal positif tidak selalu harus memiliki jarak terbesar dari solusi ideal yang negative. [18]

Metode TOPSIS mempertimbangkan kondisi dimana dengan jarak terhadap solusi ideal positif seta jarak terhadap solusi ideal negative secara bersamaan. Keluaran berupa solusi dalam metode TOPSIS didapatkan dengan cara menentukan kedekatan relative suatu alternatif pada solusi ideal positif. Proses perangkingan alternatif akan dilakukan berdasarkan prioritas nilai kedekatan suatu alternatif pada solusi ideal positif. Nantinya

daftar alternatif yang sudah diranking inilah yang akan dijadikan sebagai referensi keputusan dalam kebijakan tertentu.

Metode TOPSIS populer digunakan untuk pengambilan keputusan yang dilakukan secara praktis serta minim subjektivitas. Sebagai bagian dari alat dalam ranah sistem pendukung keputusan, TOPSIS membantu pimpinan atau manajemen menentukan kebijakan secara lebih akurat dan tepat sasaran[19]. Penggunaan metode TOPSIS menggunakan konsep yang terbilang sederhana serta mudah dipahami. Selain itu proses komputasinya efisien serta metode ini mempunyai kemampuan dalam mengukur kinerja relative dari setiap alternatif keputusan. Berikut ini tahapan metode TOPSIS melalui langkah – langkah berikut[20]

1. Membuat matrik keputusan yang sudah melewati proses normalisasi menggunakan persamaan.
2. Membuat matrik keputusan ternormalisasi dengan nilai bobot menggunakan persamaan
3. Membuat matrik dengan solusi ideal positif dan matrik solusi ideal negatif dengan menggunakan persamaan
4. Menentukan nilai preferensi untuk setiap pilihan alternatif menggunakan persamaan

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data pendaftar mahasiswa baru selama periode 2012 – 2023 di Kampus STIMIK Tunas Bangsa Banjarnegara.

3.1 Data Pemrosesan Awal

Pengolahan dataset penting dilakukan untuk memastikan tidak adanya data tidak valid karena missing value dan lainnya. Tujuannya yakni agar dataset bisa diolah dan tidak merusak hasil penelitian. Berikut ini beberapa tahapan yang dilakukan dalam pemrosesan data.

a. Pemilihan Data Awal

Untuk pemilihan kriteria data pada penelitian ini yakni meliputi variabel asal sekolah, pekerjaan orangtua, daerah asal dan program studi sebagai gambaran pilihan pendaftar.

b. Pembersihan Data (Data Cleaning)

Proses pembersihan data dilakukan untuk mencegah kesalahan pada proses asosiasi seperti adanya redundansi. Kondisi lain yakni memastikan tidak ada data hilang serta tipe data yang kurang sesuai. Jumlah data yang akan digunakan dalam penelitian ini yakni 705 data pendaftar.

c. Inisiasi Data

Inisiasi data atau perubahan bentuk ini penting dilakukan untuk memudahkan pemrosesan data mining menggunakan python dengan *tools* Google Collaboration. Ketentuan nilai inisiasi variabel penelitian bisa dilihat pada tabel 1 sedangkan hasil dari inisiasi data terlihat pada tabel 4.

Tabel 4. Data Primer Hasil Inisiasi

Mahasiswa	Asal Sekolah	Pekerjaan Orangtua	Program Studi	Daerah Asal
0	1	8	1	8
1	2	8	1	8
2	3	7	1	8
3	4	7	1	8
4	5	7	1	8

3.2 Perhitungan K-Means Clustering

Perhitungan K-Means dimulai dengan menentukan centroid awal secara acak. Berdasarkan wawancara dengan bagian pemasaran, pemilihan centroid awal yaitu nilai $k=5$. Nilai ini muncul berdasarkan jumlah lokasi promosi yang akan didatangi oleh tim penerimaan mahasiswa baru nantinya. Setelah model dilatih, selanjutnya penulis mengakses centroid dari masing – masing cluster. Pada tabel 5 menunjukkan bahwa penulis bisa mendapatkan nilai titik tengah atau centroid dari setiap cluster. Centroid ini dihitung sebagai nilai rata – rata dari semua titik data yang ada dalam semua cluster.

Tabel 5. Centroid Masing – Masing Cluster

6.67256637	6.67256637	1.54867257	1.54867257
7.41780822	4.2260274	1.53082192	7.99315068
7.33093525	6.70503597	1.58992806	7.8705036
6.82352941	6.82352941	1.54901961	3.50980392
6.13636364	4.72727273	1.42727273	5.90909091

Tahapan selanjutnya yakni memastikan apakah nilai rata – rata setiap variabel penelitian berubah atau mempunyai nilai yang berbeda pada masing – masing kategori. Pada tabel 6 di bawah terlihat hasil atau nilai rata kolom untuk masing – masing kategori.

Tabel 6. Nilai Rata – Rata Kolom Setiap Kategori

Cluster	Asal Sekolah	Pekerjaan Orangtua	Program Studi	Daerah Asal
0	6.672.566	4.159.292	1.548.673	3.477.876
1	7.417.808	4.226.027	1.530.822	7.993.151
2	7.330.935	6.705.036	1.589.928	7.870.504
3	6.823.529	6.470.588	1.549.020	3.509.804
4	6.136.364	4.727.273	1.427.273	5.909.091

Untuk mengukur seberapa baik objek ditempatkan dalam suatu kluster yang sudah dilakukan, maka peneliti mencari nilai Sillhouette Score. Dimana penghitungannya dilakukan berdasarkan jarak antara objek dengan kluster yang berdekatan. Panduan penghitungan nilai Sillhouette Score berkisar antara -1 hingga maksimal 1 dengan ketentuan semakin tinggi nilai Sillhouette Score maka hasil pengelompokan kluster dianggap semakin baik.

Untuk menghitung nilai Sillhouette Score melalui Google Collaboration menggunakan code perintah sebagai berikut.

```
silhouette_scores = silhouette_score(data2, cluster,
sample_size=len(data2), random_state=42)
silhouette_scores
```

Gambar 2. Code perintah menghitung nilai Sillhouette Score

Hasil score yang didapatkan yakni 0,4279098918854162. Berdasarkan ketentuan, nilai Sillhouette Score 0,426 termasuk dalam kategori cluster yang lemah karena berada pada rentang nilai -1. Interpretasi nilai ini masih bisa dikatakan cukup baik, namun tidak termasuk luar biasa karena masih mempunyai ruang untuk perbaikan.

3.3 Perangkingan TOPSIS

Setelah proses clustering dilakukan, tahapan selanjutnya yakni melakukan penghitungan TOPSIS untuk menganalisis keputusan bersifat multi-kriteria dan mendapatkan alternatif terbaik untuk masing – masing cluster.

a. Menentukan Keputusan Matriks Ternormalisasi

Menghitung nilai sum squared setiap kolom pada matriks, setelah itu matriks akan dinormalisasi dengan cara membaginya dengan menggunakan nilai sum squared. Pada tabel 7 terlihat matriks yang telah dinormalisasi dan siap digunakan dalam proses analisa selanjutnya.

Tabel 7. Matriks yang Telah Dinormalisasi

Cluster	Asal Sekolah	Pekerjaan Orangtua	Program Studi	Daerah Asal
0	0.432969	0.346209	0.452638	0.255583
1	0.481326	0.351764	0.447421	0.587403
2	0.475689	0.558110	0.464696	0.578390
3	0.442764	0.538595	0.452739	0.257930
4	0.398176	0.393486	0.417156	0.434249

b. Menentukan Keputusan Matriks Ternormalisasi Terbobot

Nilai bobot preferensi dari setiap kriteria dihitung berdasarkan tingkat kepentingan atau prioritas yang ditetapkan. Nilai bobot tertinggi yang digunakan pada variabel daerah asal karena patokan jarak menjadi pertimbangan penting ketika melakukan promosi. Hasil perhitungan matriks ternormalisasi terbobot terlihat melalui tabel 8 berikut.

Tabel 8. Matriks Ternormalisasi Terbobot

Cluster	Asal Sekolah	Pekerjaan Orangtua	Program Studi	Daerah Asal
0	9.238.545	8.665.295	4.418.544	19.563.100
1	8.310.380	8.528.457	4.470.068	8.512.042
2	8.408.860	5.375.287	4.303.891	8.644.687
3	9.034.153	5.570.049	4.417.554	19.385.139
4	10.045.821	7.624.162	4.794.373	11.514.130

c. Menentukan Matriks Solusi Ideal Positif dan Negatif

Dalam proses ini tujuannya yakni mencari nilai maksimum dan minimum dari kriteria asal sekolah, pekerjaan orangtua, program studi, daerah asal untuk membentuk matriks solusi ideal positif dan solusi ideal negatif.

d. Menghitung Jarak Alternatif

Performa setiap cluster dapat dievaluasi dengan cara melihat nilai hasil jarak alternatif. Misalnya untuk nilai jarak lebih kecil lebih dekat ke solusi ideal positif dan nilai jarak yang lebih besar lebih dekat dengan solusi ideal negatif, maka kondisi ini menunjukkan hasil mendekati kondisi ideal. Berlaku juga sebaliknya jika jarak dengan nilai yang lebih besar lebih dekat dengan solusi ideal positif dan nilai jarak yang lebih kecil lebih dekat dengan solusi ideal negatif. Pada tabel 9 terlihat hasil perhitungan jarak alternatif dari lima cluster. Dimana untuk kluster 0 mempunyai jarak ke solusi ideal positif 0.792940 dan jarak solusi ideal negatif 133.824663.

Tabel 9. Hasil Perhitungan Jarak Alternatif

Cluster	Positif	Negatif
0	0.792940	133.824.663
1	125.261.524	9.970.096
2	132.956.115	0.027293
3	10.777.684	118.798.934
4	65.869.884	17.322.290

e. Menggigitung Preferensi Solusi Ideal

Proses ini menghasilkan output berupa klasifikasi cluster – cluster berdasarkan seberapa dekat mereka dengan solusi ideal negatif. Jika cluster mendapatkan skor mendekati 1 artinya cluster tersebut mendekati solusi ideal negatif, sedangkan cluster yang mendekati 0 menunjukkan cluster tersebut lebih dekat dengan solusi ideal positif.

Tabel 10. Hasil Perhitungan Preferensi Solusi Ideal

Cluster	Topsis Score
0	0.994110
1	0.073726
2	0.000205
3	0.916824
4	0.208220

Hasil representasi numerik pada tabel 10 menunjukan kondisi cluster 0 memperoleh skor TOPSIS mencapai 0.994110. Interpretasi skor TOPSIS untuk cluster ini tergolong sangat tinggi karena mendekati 1. Dengan kata lain cluster ini paling dekat dengan solusi ideal negatif dalam konteks pengambilan keputusan yang diimplementasikan. Sedangkan kluster yang mendapatkan hasil TOPSIS paling rendah yakni cluster 2 dengan skor 0.000205 atau paling mendekati nilai 0 dan lebih dekat dengan solusi ideal positif.

f. Perangkingan TOPSIS

Perangkingan dilakukan dengan cara memberikan peringkat dari semua skor TOPSIS dengan ketentuan melihat dari nilai tertinggi hingga nilai terendah. Pada tabel 11 terlihat rangking 1 jatuh pada cluster 0, disusul cluster 3, cluster 4, cluster 1 dan yang paling rendah yakni cluster 2 dengan score 0.994110.

Tabel 11. Perangkingan TOPSIS

Cluster	Asal Sekolah	Pekerjaan Orangtua	Program Studi	Daerah Asal	Topsis Score	Rank
0	7	4	2	3	0.994110	1
1	7	4	2	8	0.073726	4
2	7	7	2	8	0.000205	5
3	7	6	2	4	0.916824	2
4	6	5	1	6	0.208220	3

3.4 Decision Tree

Pohon keputusan sebelumnya sudah dibuat melalui spreadsheet, agar bisa dijalankan melalui Google Collab maka perlu diunggah terlebih dahulu. Ketentuan jenis promosi dijalankan berdasarkan beberapa kriteria berikut.

- Jika jarak daerah asalnya dekat (7-8) maka dapat dilakukan promosi melalui brosur, spanduk, sosialisasi ke sekolah
- Jika jarak daerah asalnya sedang (5-6) maka dapat dilakukan promosi melalui instagram, facebook, brosur, spanduk
- Jika jarak daerah asalnya jauh (2-3) maka dapat dilakukan promosi melalui instagram, facebook, seminar, rekomendasi teman/alumni
- Jika pekerjaan orang tua adalah pedagang, petani, buruh, atau guru swasta maka dapat memberikan beasiswa
- Jika orang tua PNS maka bisa promosi melalui kerjasama Lembaga

Selanjutnya empat variabel penelitian yang terdiri dari daerah asal, asal sekolah, pekerjaan orangtua, program studi akan digabungkan menjadi satu yang akan menjadi nilai pada kolom ‘combined’. Hal ini bertujuan untuk mempermudah pemrosesan data dalam hal menentukan kode promosi. Berikut kode perintah yang digunakan untuk menggabungkan empat variabel.

3.5 Rekomendasi Promosi Untuk Masing – Masing Cluster

Tahapan selanjutnya yakni memproses data untuk mendapatkan output berupa rekomendasi jenis promosi untuk masing – masing cluster. Pertama yakni dengan menggabungkan data nilai dari empat variabel hasil K-Means dan nilai perangkingan TOPSIS menjadi satu kolom baru. Kemudian tipe data pada kolom baru tersebut dikonversi menjadi tipe data interger. Hasil datanya kemudian digabungkan dengan dataframe lain yang relevan. *Output* berupa rekomendasi jenis promosi untuk masing – masing cluster bisa terlihat Pada tabel 12 berikut.

Tabel 12. Hasil Rekomendasi Untuk Setiap Kluster

	asal_sekolah	pekerjaan_orangtua	program_studi	daerah_asal	topsis_score	rank	combine	kode_promosi
0	7	4	2	3	0.994110	1	7423	P1,P2,P6,P8,P9
1	7	4	2	8	0.073726	4	7428	P3,P4,P5,P6
2	7	7	2	8	0.000205	5	7728	P3,P4,P5,P7
3	7	6	2	4	0.916824	2	7624	P1,P2,P6,P8,P9
4	6	5	1	6	0.208220	3	6516	P1,P2,P3,P4,P9

Berikut ini interpretasi jenis promosi yang dapat dilakukan pada setiap cluster.

1. Untuk mahasiswa yang masuk ke dalam cluster 0 media promosi yang dapat digunakan yaitu instagram, facebook, beasiswa, seminar, dan rekomendasi alumni/teman
2. Untuk mahasiswa yang masuk ke dalam cluster 1 media promosi yang dapat digunakan yaitu brosur, spanduk, sosialisasi ke sekolah-sekolah, dan beasiswa
3. Untuk mahasiswa yang masuk ke dalam cluster 2 media promosi yang dapat digunakan yaitu brosur, spanduk, sosialisasi ke sekolah-sekolah, dan kerjasama lembaga
4. Untuk mahasiswa yang masuk ke dalam cluster 3 media promosi yang dapat digunakan yaitu instagram, facebook, beasiswa, seminar, dan rekomendasi alumni/teman
5. Untuk mahasiswa yang masuk ke dalam cluster 4 media promosi yang dapat digunakan yaitu instagram, facebook, brosur, spanduk, dan rekomendasi alumni/teman

4. KESIMPULAN

Pembahasan pada bagian hasil menunjukkan bahwa metode K-Means Clustering menghasilkan 5 cluster. Perhitungan Sillhouette Score mendapat 0,426 dan masuk ke dalam kategori cluster yang lemah karena berada pada rentang nilai -1. Interpretasi nilai ini masih bisa dikatakan cukup baik, namun tidak termasuk luar biasa karena masih mempunyai ruang untuk perbaikan. Pernerangan cluster menggunakan metode TOPSIS menghasilkan urutan pertama cluster yakni 0, disusul cluster 3, cluster 4, cluster 1 dan yang paling rendah yakni cluster 2 dengan score 0.994110. Hasil akhir dari penelitian ini yakni berupa rekomendasi jenis promosi untuk masing – masing cluster sesuai dengan strategi promosi yang ditentukan pada tabel pohon keputusan.

REFERENCES

- [1] S. Assylla and Nugraha, “Perancangan Strategi Pemasaran dengan Pendekatan Analisis SWOT dan Metode TOPSIS,” *J. Ris. Tek. Ind.*, pp. 129–140, 2022, doi: 10.29313/jrti.v2i2.1283.
- [2] C. L. R. Winasis, H. S. Widiandi, and B. Hadibrata, “Determinasi Keputusan Pembelian: Harga, Promosi Dan Kualitas Produk (Literature Review Manajemen Pemasaran),” *J. Ilmu Manaj. Terap.*, vol. 3, no. 4, pp. 452–462, 2022, [Online]. Available: <https://doi.org/10.31933/jemsi.v3i4>
- [3] M. T. Bolinger, M. A. Josefy, R. Stevenson, and M. A. Hitt, *Experiments in Strategy Research: A Critical Review and Future Research Opportunities*, vol. 48, no. 1. 2022. doi: 10.1177/01492063211044416.
- [4] S. B. H. Sakur, M. Silangen, and D. Tuwohingide, “Penerapan Algoritma K-Means Cluster dan Metode TOPSIS pada Pemilihan Mahasiswa kunjungan Industri,” *J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 3, pp. 851–860, 2022, [Online]. Available: <http://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/1045>
- [5] K. Khomsatun, D. Ikhsan, M. Ali, and K. Kursini, “Sistem Pengambilan Keputusan Pemilihan Lahan Tanam Di Kabupaten Wonosobo Dengan K-Means Clustering Dan Topsis,” *J. Nas. Pendidik. Tek. Inform.*, vol. 9, no. 1, p. 55, 2020, doi: 10.23887/janapati.v9i1.23073.
- [6] Khoironi, “Kombinasi Metode K-Means Dan Analytic Hierarchy Process Dalam Menentukan Penerima

- Beasiswa,” Magister Teknik Informatika, Universitas AMIKOM Yogyakarta, 2020.
- [7] Oki Oktaviarna Tensao, I Nyoman Yudi Anggara Wijaya, and Ketut Queena Fredlina, “Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara,” *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 14, no. 1, pp. 1–17, 2022, doi: 10.37424/informasi.v14i1.135.
- [8] B. T. SUTRISNO and W. Andriyani, “PENERAPAN MADM DENGAN METODE SAW UNTUK MENENTUKAN TARGET PROMOSI BERDASARKAN ASAL JURUSAN DI SEKOLAH,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 480–492, Oct. 2021, doi: 10.24176/simet.v11i2.4784.
- [9] T. Hidayat, “Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering,” *J. Sistim Inf. dan Teknol.*, pp. 19–24, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [10] Andry, Yani Maulita, and Suci Ramadani, “Sistem Pendukung Keputusan Penentuan Lokasi Promosi Penerimaan Siswa Baru Di MTS. S. Hubbul Wathan Modal Bangsa,” 2021.
- [11] M. A. Kasri and H. Jati, “Kombinasi K-Means dan Simple Additive Weighting dalam Menentukan Lokasi dan Strategi Pemasaran Universitas,” no. 2, pp. 132–141, 2020.
- [12] N. A. Rahmalinda and A. Jananto, “Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru,” *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.
- [13] I. P. D. Suarnatha, I. M. Agus, and O. Gunawan, “Jurnal Computer Science and Information Technology (CoSciTech) manusia,” *CoSciTech*, vol. 3, no. 2, pp. 73–80, 2022.
- [14] Amaliya Hani Nafisah, “Clustering Bidang Keilmuan Menggunakan Kombinasi,” vol. 04, pp. 405–413, 2023.
- [15] W. Sudrajat, I. Cholid, and J. Petrus, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer,” *J. JUPITER*, vol. 14, no. 1, pp. 27–36, 2022.
- [16] D. A. Fakhri, S. Defit, and Sumijan, “Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering,” *J. Inf. dan Teknol.*, vol. 3, pp. 160–166, 2021, doi: 10.37034/jidt.v3i3.137.
- [17] L. T. Sianturi and M. Mesran, “Penerapan Kombinasi Metode ROC dan TOPSIS Pemilihan Karyawan Terbaik Untuk Rekomendasi Promosi Jabatan,” *J. Comput. Syst. Informatics*, vol. 4, no. 1, pp. 51–60, 2022, doi: 10.47065/josyc.v4i1.2215.
- [18] Sunarti, “Perbandingan Metode TOPSIS dan SAW Untuk Pemilihan Rumah Tinggal Comparison of TOPSIS and SAW Methods For Home Selection,” *J. Inf. Syst.*, vol. 3, no. 1, p. 69, 2021.
- [19] D. Berutu, “Sistem Pendukung Keputusan Pemilihan Ibu PKK Terbaik Menggunakan Metode Fuzzy Tahani (Studi Kasus: Kantor PKK Pakpak Bharat),” *J. Comput. Syst. Informatics ...*, vol. 1, no. 4, pp. 261–268, 2020, [Online]. Available: <http://ejournal.seminar-id.com/index.php/josyc/article/view/180%0Ahttps://ejournal.seminar-id.com/index.php/josyc/article/download/180/261>
- [20] Z. Yani, D. Gusmita, and N. Pohan, “Sistem Pendukung Keputusan Penerimaan Karyawan Menggunakan Metode TOPSIS,” vol. 4307, no. June, pp. 205–210, 2022, doi: 90619941/JSSR.V205.4307.