

K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus

ORIGINALITY REPORT

13%	9%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.uin-suska.ac.id Internet Source	7%
2	Submitted to Universitas Muhammadiyah Gresik Student Paper	4%
3	Submitted to Universitas Dian Nuswantoro Student Paper	1%
4	Submitted to University of Kent at Canterbury Student Paper	<1%
5	Submitted to Universitas Putera Batam Student Paper	<1%
6	Achmad Bahauddin, Agustina Fatmawati, Febrianti Permata Sari. "ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS", Jurnal Manajemen Informatika dan Sistem Informasi, 2021 Publication	<1%
7	Aisyah Tiar Arsyad, Hanny Nurlatifah, Sunarmo Sunarmo. "Usulan Strategi Pemasaran Universitas Al Azhar Indonesia Menggunakan Metode K-Means Clustering", JURNAL AL-AZHAR INDONESIA SERI SAINS DAN TEKNOLOGI, 2024 Publication	<1%

8

Amri muliawan Nur, Muhammad Saiful2,
Hariman Bahtiar, Muhammad Taufik Hidayat.
"Penerapan Algoritma K-Means Clustering
Dalam Mengelompokkan Smartphone Yang
Rekomendasi Berdasarkan Spesifikasi",
Infotek: Jurnal Informatika dan Teknologi,
2024

Publication

<1 %

9

Fitri Nuraeni, Dede Kurniadi, Gisna Fauzian
Dermawan. "Implementation of the K-Means
Algorithm for Clustering the Characteristics of
Students Receiving Kartu Indonesia Pintar
Kuliah (KIP-K)", 2023 International Conference
on Computer Science, Information
Technology and Engineering (ICCoSITE), 2023

Publication

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

Turnitin Originality Report

Processed on: 03-Sep-2025 15:30 V

ID: 2740876349

Word Count: 4743

Submitted: 1

Similarity Index

13%

Similarity by Source

Internet Sources: 9%
Publications: 4%
Student Papers: 5%

K-Means Clustering Analysis
Untuk Menentukan Strategi
Promosi Kampus By Ukdc
Perpustakaan 2

7% match (Internet from 31-Jul-2024)

<https://ejournal.uin-suska.ac.id/index.php/jti/article/view/22492>

4% match (student papers from 29-Aug-2023)

[Submitted to Universitas Muhammadiyah Gresik on 2023-08-29](#)

1% match (student papers from 17-Mar-2023)

[Submitted to Universitas Dian Nuswantoro on 2023-03-17](#)

< 1% match (student papers from 17-Apr-2020)

[Submitted to University of Kent at Canterbury on 2020-04-17](#)

< 1% match (student papers from 01-Dec-2020)

[Submitted to Universitas Putera Batam on 2020-12-01](#)

< 1% match (Achmad Bahauddin, Agustina Fatmawati, Febrianti Permata Sari. "ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS", Jurnal Manajemen Informatika dan Sistem Informasi, 2021)

[Achmad Bahauddin, Agustina Fatmawati, Febrianti Permata Sari. "ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS", Jurnal Manajemen Informatika dan Sistem Informasi, 2021](#)

< 1% match (Aisyah Tiar Arsyad, Hanny Nurlatifah, Sunarmo Sunarmo. "Usulan Strategi Pemasaran Universitas Al Azhar Indonesia Menggunakan Metode K-Means Clustering", JURNAL AI-AZHAR INDONESIA SERI SAINS DAN TEKNOLOGI, 2024)

[Aisyah Tiar Arsyad, Hanny Nurlatifah, Sunarmo Sunarmo. "Usulan Strategi Pemasaran Universitas Al Azhar Indonesia Menggunakan Metode K-Means Clustering", JURNAL AI-AZHAR INDONESIA SERI SAINS DAN TEKNOLOGI, 2024](#)

< 1% match (Amri muliawan Nur, Muhammad Saiful2, Hariman Bahtiar, Muhammad Taufik Hidayat. "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi", Infotek: Jurnal Informatika dan Teknologi, 2024)

[Amri muliawan Nur, Muhammad Saiful2, Hariman Bahtiar, Muhammad Taufik Hidayat. "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi", Infotek: Jurnal Informatika dan Teknologi, 2024](#)

< 1% match (Fitri Nuraeni, Dede Kurniadi, Gisna Fauzian Dermawan.

"Implementation of the K-Means Algorithm for Clustering the Characteristics of Students Receiving Kartu Indonesia Pintar Kuliah (KIP-K)", 2023

International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023)

[Fitri Nuraeni, Dede Kurniadi, Gisna Fauzian Dermawan. "Implementation of the K-Means Algorithm for Clustering the Characteristics of Students Receiving Kartu Indonesia Pintar Kuliah \(KIP-K\)", 2023 International Conference on Computer Science, Information Technology and Engineering \(ICCoSITE\), 2023](#)

[K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus Agrienta](#) Bellanov1*, [Lilis Nurhayati](#)2 1,2 Jurusan Teknik Industri, Fakultas Teknik, Universitas Katolik Darma Cendika Jl. Dr. Ir. H. Soekarno No.201, Kota Surabaya, 60117 Email: agrientabellanov@ukdc.ac.id, lilisnurhayati@ukdc.ac.id ABSTRAK [Sudah banyak sektor ekonomi, pendidikan, bisnis dan lainnya yang berlomba-lomba untuk merancang strategi terbaiknya agar memperoleh target yang diinginkannya. Seperti halnya yang terjadi pada Universitas Katolik Darma Cendika. Universitas ini berupaya untuk melakukan penyusunan strategi terbaik agar semakin maju. Seperti yang diketahui bersama, kampus yang dinyatakan maju adalah kampus yang memiliki jumlah mahasiswa banyak. Hal inilah yang kemudian memicu UKDC agar terus melakukan penyusunan strategi promosi kampus. Setelah di amati lebih lanjut ternyata strategi promosi yang selama ini dilakukan oleh pihak staff Universitas masih belum maksimal, staff universitas hanya melakukan promosi di lokasi yang sama, tidak dengan melakukan pemetaan wilayah terlebih dahulu untuk melihat wilayah mana yang memiliki potensi tinggi. Berdasarkan permasalahan inilah, penulis mencoba untuk memanfaatkan data mahasiswa sejak tahun 2015, untuk diproses agar menghasilkan pemetaan wilayah yang selama ini menyumbangkan mahasiswa terbanyak, dengan membagi data menjadi 2 atribut, yakni asal sekolah dan jurusan yang dipilih di UKDC. Penelitian ini menggunakan metode algoritma K-Means Clustering dengan bantuan software Rstudio, nantinya output yang dihasilkan berupa visualisasi kluster wilayah berdasarkan 2 atribut yang ditentukan sebelumnya. Hasil yang diperoleh dari 1064 data dapat membentuk 2 kluster, dimana kluster 1 berasal dari provinsi Nusa Tenggara Timur, dan kluster 2 dari provinsi Jawa Timur. Kata kunci: K-Means Clustering, Rstudio, Mahasiswa, Strategi Promosi](#) ABSTRACT There have been many sectors of the economy, education, business and others that are competing to design the best strategy in order to get the target they want. As happened at Darma Cendika Catholic University. This university strives to carry out the best strategy preparation to be more advanced. As we all know, campuses that are declared advanced are campuses that have a large number of students. This is what then triggers UKDC to continue preparing campus promotion strategies. After further observation, it turned out that the promotion strategy that had been carried out by the university staff had not been maximized, the university staff only carried out promotions in the same location, not by mapping the area first to see which areas had high potential. Based on this problem, the author tries to utilize student data since 2015, to be processed in order to produce a mapping of areas that have contributed the most students, by dividing the data into 2 attributes, namely the origin of the school and the major chosen at UKDC. [This study uses the K-Means Clustering algorithm](#) with the help of Rstudio software, the resulting output will be a visualization of regional clusters based on 2 previously determined attributes. The results obtained from 1064 data can form 2 clusters, where cluster 1 comes from East Nusa Tenggara province, and cluster 2 from East Java province. [Keywords: K-Means Clustering, Rstudio, Students, Promotion Strategy](#) [Pendahuluan](#) Seiring dengan perjalanan [tahun](#), teknologi sistem informasi rupanya juga sudah berkembang dan mulai melekat disegala pemanfaatan berbagai bidang. Hampir seluruh kegiatan terekam menjadi banyak data yang dapat

disimpan, dengan harapan dapat dimanfaatkan kembali untuk mengetahui kejadian yang akan datang. Seperti halnya data yang terkumpul dari dunia pendidikan. Banyak sekali data yang harus disimpan untuk menjadi arsip sekolah dan sebagainya, data yang tersimpan umumnya berupa data diri siswa, asal usul sekolah siswa, bagaimana proses belajar siswa, hasil pembelajaran yang dihasilkan siswa tersebut dan masih banyak lagi. Sama halnya dengan institusi pendidikan di perguruan tinggi. Setiap tahun tenaga kependidikan harus mengolah banyak data yang berisi tentang profil mahasiswa baru. Untuk kemudian selanjutnya data-data tersebut dapat diteruskan untuk ditambah dengan nilai selama proses belajar yang dilakukan. Proses ini terus berjalan secara berulang pada institusi pendidikan, khususnya perguruan tinggi. Kumpulan data-data yang terkumpul ini akan terlihat tidak berguna apabila tidak dimanfaatkan sebaik mungkin, sebab semakin banyak data yang tersimpan, maka akan semakin susah bagi para tenaga kependidikan untuk melakukan pencarian data tersebut. Melihat dengan semakin banyaknya institusi perguruan tinggi yang hadir, maka dari itu perlu menjadi kewaspadaan khusus bagi staff tenaga kerja penerimaan mahasiswa baru, diperlukan adanya ketrampilan untuk melihat situasi dan kondisi sekitar sebagai bahan untuk menyusun strategi promosi [1], [2]. Calon mahasiswa di era sekarang ini cenderung lebih bebas untuk memilih kampus mana dan jurusan apa yang akan dipilih, seperti menurut [3]–[6] yang menyatakan bahwa minat dan bakat yang dimiliki oleh calon mahasiswa umumnya dapat menjadi salah satu faktor yang dapat mempengaruhi keputusan calon mahasiswa dan orang tua dalam memilih sekolah/ kampus mana yang akan dipilih. Universitas Katolik Darma Cendika merupakan salah satu institusi perguruan tinggi swasta yang berlokasi di Surabaya. Lokasinya cukup strategis karena berada di jalan utama merr. Namun sampai saat ini yang menjadi masalah adalah mengapa tingkat penerimaan mahasiswa baru setiap tahunnya masih tergolong rendah, padahal ada banyak pilihan jurusan yang ditawarkan. Setelah di amati lebih lanjut, rupanya strategi promosi yang selama ini dilakukan oleh pihak staff universitas masih belum maksimal, staff universitas hanya melakukan promosi dengan cara berkunjung langsung ke sekolah-sekolah untuk melakukan presentasi mengenai kampus, kemudian pembagian brosur dan sovenir. Penentuan lokasi sekolah yang dikunjungi pun setiap tahunnya sama, tidak dengan melakukan pemetaan wilayah terlebih dahulu untuk melihat wilayah mana yang memiliki potensi tinggi. Jika proses strategi ini tidak dilakukan dengan baik dan benar, tentunya akan berdampak pada penggunaan anggaran dan waktu yang kurang tepat juga, sebab strategi yang disusun secara baik akan memberikan hasil yang baik juga [7]–[10]. Berdasarkan uraian permasalahan diatas, maka dari itu dalam penelitian ini penulis mencoba untuk melakukan analisis pemilihan strategi promosi menggunakan metode K-Means Clustering. Nantinya akan diperoleh ilustrasi pemetaan wilayah persebaran mahasiswa Universitas Katolik Darma Cendika, dari pemetaan inilah strategi promosi kampus dapat disusun, agar staff dapat menentukan wilayah mana saja yang berpotensi untuk menghasilkan mahasiswa baru lebih banyak. Analisis clustering ini adalah salah satu dari banyak teknik data mining yang memiliki tujuan untuk mengelompokkan beberapa objek sesuai dengan kesamaan karakteristiknya. Sudah banyak penelitian yang menggunakan metode ini untuk menyelesaikan permasalahan serupa yakni penelitian dari [11]–[14] ada juga penelitian dari [15]–[18] yang menggunakan metode ini untuk melakukan pemetaan persebaran UMKM. Metode ini juga banyak digunakan terlebih lagi pada saat maraknya virus Covid-19 lalu, sehingga harapannya dengan pengelompokkan ini dapat diidentifikasi dimana persebaran covid terbanyak, seperti yang diteliti oleh [19]–[22]. Adapun tujuan dari penelitian ini adalah untuk mengetahui persebaran mahasiswa baru di Universitas Katolik Darma Cendika yang akan di gambarkan dalam sebuah peta area, agar lebih spesifik dalam menentukan lokasi/ target promosi kampus. Adapun manfaat dari penelitian ini adalah sebagai

petunjuk bagi staff promosi kampus dalam menentukan lokasi dan strategi promosi kampus, seperti halnya menentukan lokasi promosi berdasarkan wilayah terbanyak yang menjadi asal usul mahasiswa baru, bisa juga dengan langsung mengirimkan email ke sekolah-sekolah asal usul terbanyak yang menghasilkan mahasiswa baru di UKDC. Adapun batasan masalah yang digunakan dalam penelitian ini adalah mengolah data masa lalu dari tahun ajaran 2015/2016, 2016/2017, 2017/2018, 2018/2019, 2019/2020, 2020/2021, 2021/2022, 2022/2023. Metode Penelitian Tahapan Penelitian Agar penelitian yang dilakukan mendapatkan hasil yang maksimal, maka dari itu penulis menjabarkan langkah-langkah dalam penelitian pada flowchart dibawah ini pada gambar 1. Gambar 1. Flowchart Penelitian Heading pada level kedua dituliskan dengan boldface italics dengan menggunakan huruf besar dan huruf kecil. Heading dituliskan rata kiri. Tahap Pengumpulan Data Pada tahapan pengumpulan data ini, peneliti melakukan proses pengumpulan data di Universitas Katolik Darma Cendika Surabaya. Data yang dikumpulkan berupa data dari unit PMB (Penerimaan Mahasiswa Baru) dari tahun ajaran 2015/2016, 2016/2017, 2017/2018, 2018/2019, 2019/2020, 2020/2021, 2021/2022 dan 2022/2023. Tahap Analisis Data Metode yang digunakan dalam penelitian adalah KDD (Knowledge Discovery in Database). Metode ini merupakan bagian dari K-Means Clustering yang merupakan proses non-trivial, yakni sebuah proses untuk menemukan pola atau pemetaan yang bersifat baru, dan lebih mudah untuk dipahami [23], [24]. Adapun untuk menjalankan metode ini, terdapat 5 proses tahapan yang harus dilakukan, yakni selection, pre-processing, transformation, data training dan evaluation [25]. a. Selection (Seleksi Data) Pada tahapan ini peneliti melakukan penyortiran data, melihat dan melengkapi apabila terdapat informasi yang kurang, namun apabila informasi tidak dapat ditemukan, maka peneliti akan menghilangkan data tersebut. Seperti yang telah dijelaskan sebelumnya bahwa terdapat 4 jenis informasi/ atribut yang digunakan, dimana setiap atribut akan di kelompokkan berdasarkan informasi serupa, agar tidak keluar dari kelompok informasi yang lainnya. Adapun contoh tampilan data yang di peroleh dapat dilihat pada tabel 1 berikut ini. Tabel 1. Contoh Data No Jenis Sekolah Asal Alamat Mahasiswa Program Studi Kelamin 1 P 2 L 3 P SMK 1 Trenggalek MAN 1 Kediri SMA Negeri 1 Blitar Jl. Merdeka, no 10 Durenan Trenggalek Gang Rambutun 1, no 33, Kediri Rt 10, Rw 15, Desa Rejo, Kab Blitar Manajemen Hukum Teknik Industri b. Pre-processing Pada proses ini peneliti melakukan eliminasi pada data yang memiliki informasi tidak lengkap, dengan tujuan agar tidak merusak hasil output nantinya. c. Transformation Proses ini merupakan proses penyempurnaan data, dimana peneliti harus melakukan proses ranformasi dan konversi pada data yang diketahui belum memiliki entitas yang benar. Adapun contoh tampilan data transformasi yang telah dibagi menjadi 4 jenis adalah sebagai berikut: 1. Asal Sekolah Transformasi data pada kelompok ini merupakan pengelompokkan jenis sekolah calon mahasiswa dari kategori SLTA, yakni "SMA", "SMK", dan "MA". Adapun contoh data yang nantinya di transofrmasi dapat dilihat pada tabel 2 dibawah ini. Tabel 2. Transformasi Data Atribut Jenis Sekolah No Sekolah Asal Kategori/ Hasil Transformasi 1 SMK 1 Trenggalek SMK 2 MAN 1 Kediri MA 3 SMA Negeri 1 Blitar SMA 2. Provinsi Asal Sekolah Calon Mahasiswa Transformasi data pada atribut ini mengelompokkan provinsi asal sekolah calon mahasiswa ke dalam provinsi yang sama agar data lebih akurat. Adapun hasil transformasi pada atribut ini dapat dilihat pada tabel 4 dibawah ini. Tabel 3. Transformasi Atribut Jenis Kelamin No Sekolah Asal Kategori/ Hasil Transformasi 1 SMK 1 Trenggalek Jawa Timur 2 SMA 1 Cirebon Jawa Barat 3 SMA Negeri 1 Blitar Jawa Timur 3. Program Studi/ Jurusan Yang Dipilih Calon Mahasiswa Pada atribut ini data yang terkumpul tidak perlu dilakukan transformasi, sebab pada atribut ini hanya berisi program studi apa yang terpilih oleh calon mahasiswa, selanjutnya dari data yang terkumpul akan dilakukan konversi, pada langkah selanjutnya. Tahapan Penelitian Adapun beberapa tahapan dalam penelitian ini adalah sebagai berikut: Studi Pustaka: Penulis akan

mempelajari mengenai teori dan penerapan data mining dalam berbagai kasus, yang dapat digunakan sebagai referensi dalam penelitian ini.

Pengumpulan Data Mahasiswa: Proses pengumpulan data mahasiswa dilakukan dengan cara mengambil data dari unit kerja Penerimaan Mahasiswa Baru Universitas Katolik Darma Cendika untuk mendapatkan data yang dimaksudkan, adapun terdapat 2 atribut yang diamati, yakni: • Provinsi Asal Sekolah Calon Mahasiswa • Jurusan Yang Dipilih Seleksi Data:

Data yang diperoleh dari unit kerja PMB Universitas Katolik Darma Cendika masih berupa data mentah, sehingga perlu dilakukan transformasi data untuk menyeragamkan data dan penyeleksian data. Perhitungan [Algoritma K-Means Clustering](#): Pada penelitian ini [penulis](#) menentukan [metode k-means](#) clustering sebab [metode ini dapat](#) digunakan [untuk](#) melakukan pengelompokan data berdasarkan kesamaan jenis datanya. Hasil Clustering Dari hasil clustering ini nantinya akan muncul 2 jenis output, yakni: Kesimpulan Hasil Melakukan penarikan kesimpulan dari hasil deskriptif output software Rstudio Flowchart Penelitian Output ini akan menghadirkan hasil visual yang dapat mempermudah para pembaca untuk mengamati kelompok atau lokasi mana yang paling layak untuk ditingkatkan strategi promosinya jika dilihat dari 3 atribut yang telah ditentukan. Hasil Dan Pembahasan Persiapan Software R Studio Agar proses running pada R Studio berjalan maksimal, maka dari itu terdapat beberapa package yang harus diaktifkan terlebih dahulu untuk menunjang proses running, adapun beberapa package yang diaktifkan adalah sebagai berikut. Tabel 4. Package R Studio Yang Digunakan No Package Keterangan Package Package yang digunakan untuk mengekstrak dan memvisualisasikan 1 library(factoextra) hasil dari analisis multivariat. Library ini akan digunakan untuk visualisasi kluster dan menentukan jumlah kluster optimum Package yang dikhususkan untuk melakukan visualisasi data. Dengan 2 library(ggplot2) package ini, Anda bisa merancang grafik yang menggambarkan data melalui simbol, warna, bentuk, dan sebagainya 3 library(readr) Package yang dapat digunakan untuk membaca file, seperti file excel package yang diperlukan untuk melakukan manipulasi data frame. 4 library(dplyr) Kelebihan dplyr: Fungsional untuk eksplorasi dan transformasi data. Fungsi yang intuitif dan mudah dimengerti. 5 library(DT) Package yang dapat digunakan untuk menampilkan tabel agar lebih mudah untuk dilihat Proses Pengumpulan Data Proses pengumpulan data diperoleh dari unit PMB (Penerimaan Mahasiswa Baru). Data yang diperoleh merupakan data excel yang menunjukkan data pendaftaran mahasiswa dari semua jurusan, sejak tahun 2015 sampai tahun 2022. Terdapat beberapa data yang memiliki informasi tidak lengkap, sehingga penulis memutuskan untuk menghilangkan data tersebut agar tidak mengganggu hasil output nantinya. Adapun rekap data pendaftaran yang telah dirapikan untuk setiap tahunnya dapat dilihat pada tabel 5 berikut ini. Tabel 5. Tabel Kumpulan Data Pendaftaran Mahasiswa Jurusan 2015 2016 2017 2018 2019 2020 2021 2022 Akuntansi Manajemen Hukum Arsitektur Teknik Industri TOTAL Grand Total 59 72 50 76 29 68 74 72 63 34 26 38 22 25 7 11 14 18 12 8 14 7 4 18 6 178 205 166 196 86 1064 8 22 10 66 6 10 16 4 5 5 45 107 9 37 11 14 10 81

Sebelum menjadi seperti Tabel 5 diatas, data masih berupa data mentah sesuai dengan format dari unit PMB, dari data mentah tersebut terlihat beberapa informasi penting milik pendaftar seperti nama, jurusan, alamat rumah, sekolah asal, profesi orang tua, nama orang tua dan masih banyak lagi. Penelitian ini hanya akan membutuhkan data jurusan, jenis sekolah terakhir (SMA/ SMK) dan provinsi asal sekolah, penulis memutuskan untuk memvisualisasikan provinsi dari pada kota asal, sebab untuk standar jumlah pendaftar yang masih rendah, apabila pemetaan dilakukan berdasarkan kota, dipastikan visualisasi/ pemetaan tidak akan terbentuk sempurna, maka dari itu penulis menggunakan provinsi sebagai dasar visualisasi. Tabel 6 dibawah ini akan menunjukkan contoh tampilan data mentah yang akan menjadi input di R Studio. Jurusan Akuntansi Akuntansi Akuntansi Manajemen

Manajemen Manajemen Hukum Hukum Hukum Arsitektur Arsitektur Industri Industri Tabel 6. Contoh Tampilan Data Mentah Asal Sekolah SMA HANG TUAH 2 SMA SENOPATI SMAS ST MARKUS RUTENG SMAN 2 BANJARBARU SMAK KARITAS 3 SMA MUHAMMADIYAH 3 SMA ST. BONAVENTURA SMA KR. PETRA 2 SMAN 1 BLITAR SMA SANTA MARIA 3 SMI SMA BUNDA HATI KUDUS SMK ST. LOUIS Jenis Kelompok Kota Provinsi SMA 1 SIDOARJO JAWA TIMUR SMA 1 SIDOARJO JAWA TIMUR SMA 1 MANGGARAI NUSA TENGGARA TIMUR SMA 1 BANJAR KALIMANTAN BARU SELATAN SMA 1 SURABAYA JAWA TIMUR SMA 1 SURABAYA JAWA TIMUR SMA 1 MADIUN JAWA TIMUR SMA 1 SURABAYA JAWA TIMUR SMA 1 BLITAR JAWA TIMUR SMA 1 CIMAH I JAWA BARAT SMI 3 SURABAYA JAWA TIMUR SMA 1 SURABAYA JAWA TIMUR SMK 2 SURABAYA JAWA TIMUR

Pada Tabel 6 diatas dapat dilihat bahwa pengelompokan sudah dilakukan pada jenis sekolah asal, dimana skor 1 adalah SMA, skor 2 adalah SMK, dan skor 3 adalah lainnya. Kemudian untuk pengelompokan provinsi penulis telah merekap dari keseluruhan data yang berjumlah 1064 data, nantinya provinsi itu yang akan digunakan, untuk provinsi sumatra utara, sumatera selatan akan digabungkan menjadi provinsi Sumatera, karena data pendaftar yang berasal dari provinsi tersebut hanya 1-3 pendaftar saja, sama halnya dengan kalimantan utara, kalimantan barat, akan digabungkan menjadi provinsi kalimantan saja, dan seterusnya, adapun rekap seluruh provinsi dari 1064 data (data 2015 – 2022) dapat dilihat pada Tabel 7 berikut.

Tabel 7. Rekap Data Provinsi Yang Digunakan Jurusan Akuntansi Manajemen Hukum Industri Arsitektur Bali Jawa Barat Jakarta Ntt Sumatra Kaltim Jawa Tengah Kalimantan Papua Sulawesi Timor Leste Jawa Timur Maluku Total Grand Total

2	2	2	1	0	1	24	30	5	10	1	3	5	8	0	2	0
5	6	6	0	0	282	357	0	1	327	426	0	1	0	16	10	1
0	2	3	1	0	111	0	145	1064	1	0	0	3	3	0	1	0
0	0	3	0	86	0	97	1	1	0	5	3	3	0	0	0	1
55	0	69														

Setelah merapikan data pada excel, maka langkah selanjutnya adalah masuk ke software R Studio untuk memulai melakukan tahap running. Untuk penulisan syntax pada lembar kerja R Studio dapat dilihat dibawah ini.

```

data <- rekap data View(rekap) #melihat data summary(rekap)
#ringkasan data attach(rekap) #memisahkan variabel index<-data kluster
<- index[2:6] a <- scale(kluster) jarak <- get_dist(a) fviz_nbclust(kluster,
kmeans, method = "wss") fviz_nbclust(kluster, kmeans, method =
"silhouette") fviz_dist(jarak, gradient=list(low="#00AFBB", mid="white",
high="#FC4E07")) k3 <- kmeans(a, centers = 2, nstart = 10) #penentuan
jumlah cluster k3 fviz_cluster(k3, data=kluster) #visualisasi hasil
pengelompokan table(k3$kluster, rekap$jurusan) final=data.frame(rekap,
klaster$cluster) final Interpretasi Output/ Hasil Setelah seluruh syntax
yang ada dimasukkan/ dituliskan pada lembar kerja R Studio, maka
langkah selanjutnya adalah melakukan interpretasi dari hasil running yang
dihasilkan. Terdapat 3 output berupa gambar/visualisasi dari data yang
telah di running. Output yang pertama dengan syntax:
fviz_nbclust(kluster, kmeans, method = "wss") fviz_nbclust(kluster,
kmeans, method = "silhouette") Menghasilkan Gambar 2 seperti dibawah
ini, dimana output ini merupakan grafik yang akan menunjukkan berapa
jumlah optimal kluster yang terbentuk dari persebaran data pendaftaran
yang diolah. Gambar grafik yang pertama menggunakan metode wss,
dimana dengan metode ini terlihat grafik mulai turun drastis pada angka 2,
hal ini berarti bahwa jumlah kluster yang optimal dibentuk adalah 2 kluster
saja. kemudian untuk gambar grafik yang kedua adalah menggunakan
metode silhouette, hasil ang ditunjukkan menggunakan metode ini adalah
2 kluster, terlihat ada garis putus-putus yang berada pada angka 2, yang
berarti bahwa kluster optimal yang terbentuk adalah 2 saja. Secara umum
dapat dikatakan bahwa persebaran data pendaftaran mahasiswa memiliki
2 kelompok data yang berbeda, dimana setiap kelompok merupakan data
yang memiliki kesamaan dalam hal persebaran datanya. Gambar 2. Grafik
Jumlah Optimal Kluster Output selanjutnya adalah grafik Heat Mape yang
dapat dimunculkan dengan menuliskan syntax seperti dibawah ini.
fviz_dist(jarak, gradient=list(low="#00AFBB", mid="white",

```

high="#FC4E07")) Grafik Heat Mape merupakan output yang menampilkan visualisasi persebaran data dengan menggunakan warna. Warna merah menunjukkan persebaran datanya semakin jauh, sedangkan jika warnanya semakin hijau maka persebaran datanya semakin dekat. Untuk mengetahui data apa saja yang saling dekat dan jauh dapat dilihat pada garis kordinat X dan Y nya, dimana terdapat angka-angka yang menempel pada garis, angka tersebut merupakan informasi data ke... (contoh: data ke 7, data ke 13, dan sebagainya). Adapun tampilan grafik heat mape dapat dilihat pada Gambar 3 dibawah ini. Gambar 3. Grafik Heat Mape Syntax dibawah ini akan menampilkan visualisasi pemetaan persebaran data pendaftar UKDC yang dibedakan dari beberapa provinsi dan jurusan, adapun syntax yang digunakan adalah sebagai berikut. `k3 <- kmeans(a, centers = 2, nstart = 10) #penentuan jumlah cluster k3 fviz_cluster(k3, data=kluster) #visualisasi hasil pengelompokan` Mengikuti saran yang muncul pada Gambar 3 diatas, bahwa jumlah kluster yang optimal pada data ini adalah 2, maka penulis juga menuliskan 2 kluster pada syntax. Dari hasil yang diperoleh dapat dilihat bahwa cluster 1 diwakili oleh warna merah, kemudian kluster yang kedua diwakili oleh warna biru. Terdapat 1 data saja yang masuk pada kluster 2 yakni data nomor 12, dan data yang lainnya masuk pada kluster 1. Hal ini berarti persebaran data pada data nomor 12 terlalu jauh jika dibandingkan dengan data lainnya. Untuk lebih jelasnya dapat dilihat pada Gambar 5 dibawah ini, kemudian untuk gambar plot kluster dapat dilihat pada Gambar 4 dibawah ini. Gambar 4. Plot Kluster Syntax yang terakhir menunjukkan hasil dalam tabel, untuk lebih memudahkan pembaca dalam melakukan interpretasi data, adapun syntax yang digunakan adalah seperti dibawah ini, kemudian untuk tabel hasil running ada pada gambar 5. `table(k3$kluster, rekap$jurusan) final=data.frame(rekap, klaster$cluster) final` Gambar 5. Tabel Hasil Running Pada gambar diatas terlihat bahwa kluster 1 dianggotai oleh 12 data, dengan pendaftar tertinggi selama tahun 2015- 2022 berasal dari Nusa Tenggara Timur dengan jumlah mahasiswa akuntansi 24 orang, manajemen 30 orang, hukum 16 orang, teknik industri 3 orang, dan arsitektur 5 orang. 12 provinsi ini memiliki kesamaan persebaran data maka dari itu seluruhnya masuk ke dalam kluster 1. Sedangkan untuk kluster 2 hanya di isi oleh data nomor 12 yakni provinsi jawa timur. Seperti yang dapat dilihat jumlah pendaftar yang berasal dari jawa timur sangat tinggi, sehingga terlihat jauh sekali perbedaannya dengan provinsi lain, tidak heran jika provinsi jawa timur menjadi kluster tertinggi atau kluster sendiri. secara umum jika dilihat dari data, pendaftar yang berasal dari jawa timur didominasi oleh kota Surabaya dan Sidoarjo. Simpulan Setelah melakukan analisa diatas maka penulis dapat menarik kesimpulan bahwa total 1064 data pendaftaran mahasiswa baru sejak tahun 2015- 2022 yang dianalisis menggunakan metode K-Means Clustering menghasilkan 2 kluster yang optimal, dimana setiap kluster pasti memiliki persebaran data yang hampir mirip maka dari itu mereka dikelompokkan menjadi 1 kluster yang sama. Kluster 1 dianggotai oleh 12 data (provinsi), dengan provinsi pendaftar tertinggi berasal dari Nusa Tenggara Timur dengan total 78 mahasiswa, provinsi Nusa Tenggara ini mencakup Kabupaten Flores, Kabupaten Ende dan Kabupaten Manggarai. Sementara untuk kluster 2 merupakan cluster yang memiliki jumlah data paling banyak dibandingkan dengan kluster 1, dan terlihat adanya selisih perbedaan data yang cukup jauh antara kluster 1 dengan kluster 2. Kluster 2 hanya memiliki 1 anggota saja yakni provinsi Jawa Timur (Kota Surabaya, Kabupaten Sidoarjo, dan sekitarnya), dengan total data pendaftar sebanyak 891 mahasiswa, dimana dalam provinsi ini Kota Surabaya menjadi kota yang paling banyak menyumbang mahasiswa. Dengan adanya hasil ini diharapkan unit PMB dapat lebih menekan upaya promosi kampus ke Provinsi Jawa Timur khususnya Kota Surabaya, dan Provinsi Nusa Tenggara Timur, misalnya dengan cara menitipkan brosur dengan sistem relasi kepada mahasiswa untuk dibawa pada saat pulang kampung, atau sedang berkunjung ke sekolah SMA/ SMK sebelumnya. Untuk memperluas hasil analisis penelitian

ini, maka dari itu penulis selanjutnya dapat melakukan penelitian yang bertujuan untuk melakukan analisis kluster yang lebih spesifik yakni dengan melakukan pengelompokan data berdasarkan daerah (kabupaten/kota). Daftar Pustaka [1] A. Wardhana, "Strategi Promosi Penerimaan Siswa Baru (Studi Kasus Smk Kesehatan Cipta Bhakti Husada Yogyakarta)," CHANNEL J. Komun., vol. 6, no. 1, p. 96, 2018, doi: 10.12928/channel.v6i1.10215. [2] F. I. Nurlita, Analisis SWOT Sebagai Strategi Meningkatkan Daya Saing Usaha Kerajinan Sapu Glagah. 2020. [3] T. Triyono, D. Dasmadi, and A. F. A. TNK, "Pengaruh Promosi, Biaya, Fasilitas, Akreditasi, Dan Lokasi Universitas Boyolali Terhadap Minat Calon Mahasiswa Baru," EKOBIS J. Ilmu Manaj. dan Akunt., vol. 9, no. 2, pp. 220–229, 2021. [4] Y. M. LAROSA, "Analisis Faktor-Faktor Yang Mempengaruhi Motivasi Mahasiswa Dalam Memilih Konsentrasi Jurusan Manajemen Di Stie Pembnas Nias," J. Akunt. dan Manaj. Pembnas, vol. 8, no. 1, pp. 53–62, 2021. [5] E. D. S. Mulyani, C. R. Hidayat, and T. C. Ulfa, "Sistem Pakar Untuk Menentukan Jurusan Kuliah Berdasarkan Minat dan Bakat Siswa SMA Dengan Menggunakan Metode Forward Chaining," CSRID (Computer Sci. Res. Its Dev. Journal), vol. 10, no. 2, p. 80, 2021, doi: 10.22303/csrid.10.2.2018.80-92. [6] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," J. Sains Teknol. Transp. Marit., vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30. [7] R. D. Dana, C. L. Rohmat, and A. R. Rinaldi, "Strategi Marketing Penerimaan Mahasiswa Baru Menggunakan Machine Learning dengan Teknik Clustering," J. Pengemb. IT, vol. 04, no. 2, pp. 201–204, 2018, doi: 10.30591/jpit.v4i2-2.1879. [8] T. Tukino and B. Huda, "Penerapan Algoritma K-Means Untuk Mendukung Keputusan Dalam Pemilihan Tema Tugas Akhir Pada Prodi Sistem Informasi Universitas Buana Perjuangan Karawang.," Techno Xplore J. Ilmu Komput. dan Teknol. Inf., vol. 4, no. 1, pp. 1–10, 2019, doi: 10.36805/technoxplore.v4i1.542. [9] Y. Firmansyah and D. Purwaningtias, "725-2482-2-Pb," vol. 01, no. 02, pp. 70–82, 2017. [10] E. Nurhayati, "Strategi Peningkatan Produktivitas Untuk Mencapai Target Produktivitas Dan Efisiensi Perusahaan," Ind. Eng. J. Univ. Sarjanawiyata Tamansiswa, vol. 2, no. 1, pp. 62–69, 2018. [11] I. Irmayansyah and S. E. Triyono, "Penerapan Algoritma K-Means Untuk Pemetaan Potensi Calon Mahasiswa Baru," Teknois J. Ilm. Teknol. Inf. dan Sains, vol. 12, no. 2, pp. 139–150, 2022, doi: 10.36350/jbs.v12i2.139. [12] R. Budiman and R. Anto, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)," ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform., vol. 6, no. 1, p. 6, 2019, doi: 10.30656/protekinfo.v6i1.1691. [13] L. Hakim, "Universitas Xyz Menggunakan Algoritma K-Means Clustering," J. Sist. Inf. Dan Teknologi, vol. 2, pp. 80–86, 2019. [14] H. Hairani, D. Susilowati, I. Puji Lestari, K. Marzuki, and L. Z. A. Mardedi, "Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 2, pp. 275–282, 2022, doi: 10.30812/matrik.v21i2.1542. [15] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," JOINS (Journal Inf. Syst., vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469. [16] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," JOINS (Journal Inf. Syst., vol. 6, no. 2, pp. 211–220, 2021, doi: 10.33633/joins.v6i2.5310. [17] T. Zulyanti and Noeryanti, "Perbandingan Pengelompokan Usaha Mikro Kecil dan Menengah di Kabupaten Klaten Tahun 2019 dengan Metode K-Means dan Clustering Large Application," J. Stat. Ind. dan Komputasi, vol. 7, no. 1, pp. 46–59, 2022. [18] Y. Ansori and C. Wulandari, "CRISP-DM Method On Indonesian Micro Industries (UMKM) Using K- Means Clustering Algorithm," MATICS J. Ilmu Komput. dan Teknol. Inf. (Journal Comput. Sci. Inf. Technol., vol. 14, no. 2, pp. 35–40, 2022, doi: 10.18860/mat.v14i2.13760. [19] Z. Nabila,

K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus

by Ukdc Perpustakaan 2

Submission date: 03-Sep-2025 03:30PM (UTC+0700)

Submission ID: 2740876349

File name: 22492-67225-1-PB_3.pdf (245.49K)

Word count: 4743

Character count: 26787

1
K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus6
Agrianta Belanov^{1*}, Lilis Nurhayati²12
Jurusan Teknik Industri, Fakultas Teknik, Universitas Katolik Dharma Cendika

Jl. Dr. Ir. H. Soekarno No.201, Kota Surabaya, 60117

Email: agriantabelanov@ukdc.ac.id, lilisnurhayati@ukdc.ac.id**ABSTRAK**

1
Sudah banyak sektor ekonomi, pendidikan, bisnis dan lainnya yang berlomba-lomba untuk merancang strategi terbaiknya agar memperoleh target yang diinginkan. Seperti halnya yang terjadi pada Universitas Katolik Dharma Cendika. Universitas ini berupaya untuk melakukan penyusunan strategi terbaik agar semakin maju. Seperti yang diketahui bersama, kampus yang dinyatakan maju adalah kampus yang memiliki jumlah mahasiswa banyak. Hal inilah yang kemudian memicu UKDC agar terus melakukan penyusunan strategi promosi kampus. Setelah di amati lebih lanjut ternyata strategi promosi yang selama ini dilakukan oleh pihak staff Universitas masih belum maksimal, staff universitas hanya melakukan promosi di lokasi yang sama, tidak dengan melakukan pemetaan wilayah terlebih dahulu untuk melihat wilayah mana yang memiliki potensi tinggi. Berdasarkan permasalahan inilah, penulis mencoba untuk memanfaatkan data mahasiswa sejak tahun 2015, untuk diproses agar menghasilkan pemetaan wilayah yang selama ini menyumbangkan mahasiswa terbanyak, dengan membagi data menjadi 2 atribut, yakni asal sekolah dan jurusan yang dipilih di UKDC. Penelitian ini menggunakan metode algoritma *K-Means Clustering* dengan bantuan software Rstudio, nantinya output yang dihasilkan berupa visualisasi kluster wilayah berdasarkan 2 atribut yang ditentukan sebelumnya. Hasil yang diperoleh dari 1064 data dapat membentuk 2 kluster, dimana kluster 1 berasal dari provinsi Nusa Tenggara Timur, dan kluster 2 dari provinsi Jawa Timur.

Kata kunci: K-Means Clustering, Rstudio, Mahasiswa, Strategi Promosi**ABSTRACT**

There have been many sectors of the economy, education, business and others that are competing to design the best strategy in order to get the target they want. As happened at Dharma Cendika Catholic University. This university strives to carry out the best strategy preparation to be more advanced. As we all know, campuses that are declared advanced are campuses that have a large number of students. This is what then triggers UKDC to continue preparing campus promotion strategies. After further observation, it turned out that the promotion strategy that had been carried out by the university staff had not been maximized, the university staff only carried out promotions in the same location, not by mapping the area first to see which areas had high potential. Based on this problem, the author tries to utilize student data since 2015, to be processed in order to produce a mapping of areas that have contributed the most students, by dividing the data into 2 attributes, namely the origin of the school and the major chosen at UKDC. This study uses the *K-Means Clustering* algorithm with the help of Rstudio software, the resulting output will be a visualization of regional clusters based on 2 previously determined attributes. The results obtained from 1064 data can form 2 clusters, where cluster 1 comes from East Nusa Tenggara province, and cluster 2 from East Java province.

Keywords: *K-Means Clustering, Rstudio, Students, Promotion Strategy***Pendahuluan**

Seiring dengan perjalanan tahun, teknologi sistem informasi rupanya juga sudah berkembang dan mulai melekat disegala pemanfaatan berbagai bidang. Hampir seluruh kegiatan terekam menjadi banyak data yang dapat disimpan, dengan harapan dapat dimanfaatkan kembali untuk mengetahui kejadian yang akan datang. Seperti halnya data yang terkumpul dari dunia pendidikan. Banyak sekali data yang harus disimpan untuk menjadi arsip sekolah dan sebagainya, data yang tersimpan umumnya berupa data diri siswa, asal usul sekolah siswa, bagaimana proses belajar siswa, hasil pembelajaran yang dihasilkan siswa tersebut dan masih banyak lagi.

Sama halnya dengan institusi pendidikan di perguruan tinggi. Setiap tahun tenaga kependidikan harus mengolah banyak data yang berisi tentang profil mahasiswa baru. Untuk kemudian selanjutnya data-data tersebut dapat diteruskan untuk ditambah dengan nilai selama proses belajar yang dilakukan. Proses ini terus berjalan secara berulang pada institusi pendidikan, khususnya perguruan tinggi. Kumpulan data-data yang terkumpul ini akan terlihat tidak berguna apabila tidak dimanfaatkan sebaik mungkin, sebab semakin banyak data yang tersimpan, maka akan semakin susah bagi para tenaga kependidikan untuk melakukan pencarian data tersebut.



Melihat dengan semakin banyaknya institusi perguruan tinggi yang hadir, maka dari itu perlu menjadi kewaspadaan khusus bagi staff tenaga kerja penerimaan mahasiswa baru, diperlukan adanya ketrampilan untuk melihat situasi dan kondisi sekitar sebagai bahan untuk menyusun strategi promosi [1], [2]. Calon mahasiswa di era sekarang ini cenderung lebih bebas untuk memilih kampus mana dan jurusan apa yang akan dipilih, seperti menurut [3]-[6] yang menyatakan bahwa minat dan bakat yang dimiliki oleh calon mahasiswa umumnya dapat menjadi salah satu faktor yang dapat mempengaruhi keputusan calon mahasiswa dan orang tua dalam memilih sekolah/ kampus mana yang akan dipilih.

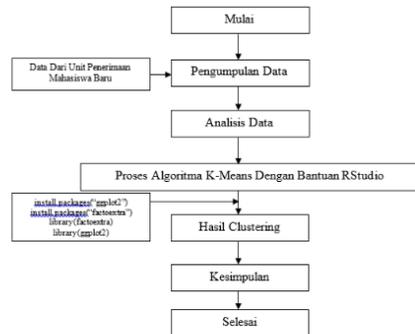
Universitas Katolik Dharma Cendika merupakan salah satu institusi perguruan tinggi swasta yang berlokasi di Surabaya. Lokasinya cukup strategis karena berada di jalan utama merr. Namun sampai saat ini yang menjadi masalah adalah mengapa tingkat penerimaan mahasiswa baru setiap tahunnya masih tergolong rendah, padahal ada banyak pilihan jurusan yang ditawarkan. Setelah di amati lebih lanjut, rupanya strategi promosi yang selama ini dilakukan oleh pihak staff universitas masih belum maksimal, staff universitas hanya melakukan promosi dengan cara berkunjung langsung ke sekolah-sekolah untuk melakukan presentasi mengenai kampus, kemudian pembagian brosur dan souvenir. Penentuan lokasi sekolah yang dikunjungi pun setiap tahunnya sama, tidak dengan melakukan pemetaan wilayah terlebih dahulu untuk melihat wilayah mana yang memiliki potensi tinggi. Jika proses strategi ini tidak dilakukan dengan baik dan benar, tentunya akan berdampak pada penggunaan anggaran dan waktu yang kurang tepat juga, sebab strategi yang disusun secara baik akan memberikan hasil yang baik juga [7]-[10].

Berdasarkan uraian permasalahan diatas, maka dari itu dalam penelitian ini penulis mencoba untuk melakukan analisis pemilihan strategi promosi menggunakan metode *K-Means Clustering*. Nantinya akan diperoleh ilustrasi pemetaan wilayah persebaran mahasiswa Universitas Katolik Dharma Cendika, dari pemetaan inilah strategi promosi kampus dapat disusun, agar staff dapat menentukan wilayah mana saja yang berpotensi untuk menghasilkan mahasiswa baru lebih banyak. Analisis clustering ini adalah salah satu dari banyak teknik data mining yang memiliki tujuan untuk mengelompokkan beberapa objek sesuai dengan kesamaan karakternya. Sudah banyak penelitian yang menggunakan metode ini untuk menyelesaikan permasalahan serupa yakni penelitian dari [11]-[14] ada juga penelitian dari [15]-[18] yang menggunakan metode ini untuk melakukan pemetaan persebaran UMKM. Metode ini juga banyak digunakan terlebih lagi pada saat maraknya virus Covid-19 lalu, sehingga harapannya dengan pengelompokan ini dapat diidentifikasi dimana persebaran covid terbanyak, seperti yang diteliti oleh [19]-[22]. Adapun tujuan dari penelitian ini adalah untuk mengetahui persebaran mahasiswa baru di Universitas Katolik Dharma Cendika yang akan di gambarkan dalam sebuah peta area, agar lebih spesifik dalam menentukan lokasi/ target promosi kampus. Adapun manfaat dari penelitian ini adalah sebagai petunjuk bagi staff promosi kampus dalam menentukan lokasi dan strategi promosi kampus, seperti halnya menentukan lokasi promosi berdasarkan wilayah terbanyak yang menjadi asal usul mahasiswa baru, bisa juga dengan langsung mengirimkan email ke sekolah-sekolah asal usul terbanyak yang menghasilkan mahasiswa baru di UKDC. Adapun batasan masalah yang digunakan dalam penelitian ini adalah mengolah data masa lalu dari tahun ajaran 2015/2016, 2016/2017, 2017/2018, 2018/2019, 2019/2020, 2020/2021, 2021/2022, 2022/2023.

Metode Penelitian

Tahapan Penelitian

Agar penelitian yang dilakukan mendapatkan hasil yang maksimal, maka dari itu penulis menjabarkan langkah-langkah dalam penelitian pada flowchart dibawah ini pada gambar 1.



Gambar 1. Flowchart Penelitian



3
Heading pada level kedua dituliskan dengan *boldface italics* dengan menggunakan huruf besar dan huruf kecil. Heading dituliskan rata kiri.

Tahap Pengumpulan Data

Pada tahapan pengumpulan data ini, peneliti melakukan proses pengumpulan data di Universitas Katolik Darma Cendika Surabaya. Data yang dikumpulkan berupa data dari unit PMB (Penerimaan Mahasiswa Baru) dari tahun ajaran 2015/2016, 2016/2017, 2017/2018, 2018/2019, 2019/2020, 2020/2021, 2021/2022 dan 2022/2023.

Tahap Analisis Data

Metode yang digunakan dalam penelitian adalah KDD (*Knowledge Discovery in Database*). Metode ini merupakan bagian dari *K-Means Clustering* yang merupakan proses non-trivial, yakni sebuah proses untuk menemukan pola atau pemetaan yang bersifat baru, dan lebih mudah untuk dipahami [23], [24]. Adapun untuk menjalankan metode ini, terdapat 5 proses tahapan yang harus dilakukan, yakni selection, pre-processing, transformation, data training dan evaluation [25].

a. *Selection* (Seleksi Data)

Pada tahapan ini peneliti melakukan penyortiran data, melihat dan melengkapi apabila terdapat informasi yang kurang, namun apabila informasi tidak dapat ditemukan, maka peneliti akan menghilangkan data tersebut. Seperti yang telah dijelaskan sebelumnya bahwa terdapat 4 jenis informasi/ atribut yang digunakan, dimana setiap atribut akan di kelompokkan berdasarkan informasi serupa, agar tidak keluar dari kelompok informasi yang lainnya. Adapun contoh tampilan data yang di peroleh dapat dilihat pada tabel 1 berikut ini.

Tabel 1. Contoh Data

No	Jenis Kelamin	Sekolah Asal	Alamat Mahasiswa	Program Studi
1	P	SMK 1 Trenggalek	Jl. Merdeka, no 10 Durenan Trenggalek	Manajemen
2	L	MAN 1 Kediri	Gang Rambutan 1, no 33, Kediri	Hukum
3	P	SMA Negeri 1 Blitar	Rt 10, Rw 15, Desa Rejo, Kab Blitar	Teknik Industri

b. *Pre-processing*

Pada proses ini peneliti melakukan eliminasi pada data yang memiliki informasi tidak lengkap, dengan tujuan agar tidak merusak hasil output nantinya.

c. *Transformation*

Proses ini merupakan proses penyempurnaan data, dimana peneliti harus melakukan proses ranformasi dan konversi pada data yang diketahui belum memiliki entitas yang benar. Adapun contoh tampilan data transformasi yang telah dibagi menjadi 4 jenis adalah sebagai berikut:

1. Asal Sekolah

Transformasi data pada kelompok ini merupakan pengelompokkan jenis sekolah calon mahasiswa dari kategori SLTA, yakni "SMA", "SMK", dan "MA". Adapun contoh data yang nantinya di transformasi dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Transformasi Data Atribut Jenis Sekolah

No	Sekolah Asal	Kategori/ Hasil Transformasi
1	SMK 1 Trenggalek	SMK
2	MAN 1 Kediri	MA
3	SMA Negeri 1 Blitar	SMA

2. Provinsi Asal Sekolah Calon Mahasiswa

Transformasi data pada atribut ini mengelompokkan provinsi asal sekolah calon mahasiswa ke dalam provinsi yang sama agar data lebih akurat. Adapun hasil transformasi pada atribut ini dapat dilihat pada tabel 4 dibawah ini.

Tabel 3. Transformasi Atribut Jenis Kelamin

No	Sekolah Asal	Kategori/ Hasil Transformasi
1	SMK 1 Trenggalek	Jawa Timur



2	SMA 1 Cirebon	Jawa Barat
3	SMA Negeri 1 Blitar	Jawa Timur

3. Program Studi/ Jurusan Yang Dipilih Calon Mahasiswa
 Pada atribut ini data yang terkumpul tidak perlu dilakukan transformasi, sebab pada atribut ini hanya berisi program studi apa yang terpilih oleh calon mahasiswa, selanjutnya dari data yang terkumpul akan dilakukan konversi, pada langkah selanjutnya.

Tahapan Penelitian

Adapun beberapa tahapan dalam penelitian ini adalah sebagai berikut:

Studi Pustaka:

Penulis akan mempelajari mengenai teori dan penerapan data mining dalam berbagai kasus, yang dapat digunakan sebagai referensi dalam penelitian ini.

Pengumpulan Data Mahasiswa:

Proses pengumpulan data mahasiswa dilakukan dengan cara mengambil data dari unit kerja Penerimaan Mahasiswa Baru Universitas Katolik Dharma Cendika untuk mendapatkan data yang dimaksudkan, adapun terdapat 2 atribut yang diamati, yakni:

- Provinsi Asal Sekolah Calon Mahasiswa
- Jurusan Yang Dipilih

Seleksi Data:

Data yang diperoleh dari unit kerja PMB Universitas Katolik Dharma Cendika masih berupa data mentah, sehingga perlu dilakukan transformasi data untuk menyeragamkan data dan penyeleksian data.

Perhitungan Algoritma K-Means Clustering:

Pada penelitian ini penulis menentukan metode *k-means clustering* sebab metode ini dapat digunakan untuk melakukan pengelompokan data berdasarkan kesamaan jenis datanya.

Hasil Clustering

Dari hasil clustering ini nantinya akan muncul 2 jenis output, yakni:
 Kesimpulan Hasil
 Melakukan penarikan kesimpulan dari hasil deskriptif output software Rstudio
 Flowchart Penelitian
 Output ini akan menghadirkan hasil visual yang dapat mempermudah para pembaca untuk mengamati kelompok atau lokasi mana yang paling layak untuk ditingkatkan strategi promosinya jika dilihat dari 3 atribut yang telah ditentukan.

Hasil Dan Pembahasan

Persiapan Software R Studio

Agar proses running pada R Studio berjalan maksimal, maka dari itu terdapat beberapa package yang harus diaktifkan terlebih dahulu untuk menunjang proses running, adapun beberapa package yang diaktifkan adalah sebagai berikut.

Tabel 4. Package R Studio Yang Digunakan

No	Package	Keterangan Package
1	library(factoextra)	Package yang digunakan untuk mengekstrak dan memvisualisasikan hasil dari analisis multivariat. Library ini akan digunakan untuk visualisasi kluster dan menentukan jumlah kluster optimum
2	library(ggplot2)	Package yang dikhususkan untuk melakukan visualisasi data. Dengan package ini, Anda bisa merancang grafik yang menggambarkan data melalui simbol, warna, bentuk, dan sebagainya
3	library(readr)	Package yang dapat digunakan untuk membaca file, seperti file excel
4	library(dplyr)	Kelebihan dplyr: Fungsional untuk eksplorasi dan transformasi data. Fungsi yang intuitif dan mudah dimengerti.



5 library(DT) Package yang dapat digunakan untuk menampilkan tabel agar lebih mudah untuk dilihat

Proses Pengumpulan Data

Proses pengumpulan data diperoleh dari unit PMB (Penerimaan Mahasiswa Baru). Data yang diperoleh merupakan data excel yang menunjukkan data pendaftaran mahasiswa dari semua jurusan, sejak tahun 2015 sampai tahun 2022. Terdapat beberapa data yang memiliki informasi tidak lengkap, sehingga penulis memutuskan untuk menghilangkan data tersebut agar tidak mengganggu hasil output nantinya. Adapun rekap data pendaftaran yang telah dirapikan untuk setiap tahunnya dapat dilihat pada tabel 5 berikut ini.

Tabel 5. Tabel Kumpulan Data Pendaftaran Mahasiswa

Jurusan	2015	2016	2017	2018	2019	2020	2021	2022
Akuntansi	59	72	50	76	29	8	22	9
Manajemen	68	74	72	63	34	10	66	37
Hukum	26	38	22	25	7	6	10	11
Arsitektur	11	14	18	12	8	16	4	14
Teknik Industri	14	7	4	18	6	5	5	10
TOTAL	178	205	166	196	86	45	107	81
Grand Total	1064							

Sebelum menjadi seperti Tabel 5 diatas, data masih berupa data mentah sesuai dengan format dari unit PMB, dari data mentah tersebut terlihat beberapa informasi penting milik pendaftar seperti nama, jurusan, alamat rumah, sekolah asal, profesi orang tua, nama orang tua dan masih banyak lagi. Penelitian ini hanya akan membutuhkan data jurusan, jenis sekolah terakhir (SMA/ SMK) dan provinsi asal sekolah, penulis memutuskan untuk memvisualisasikan provinsi dari pada kota asal, sebab untuk standar jumlah pendaftar yang masih rendah, apabila pemetaan dilakukan berdasarkan kota, dipastikan visualisasi/ pemetaan tidak akan terbentuk sempurna, maka dari itu penulis menggunakan provinsi sebagai dasar visualisasi. Tabel 6 dibawah ini akan menunjukkan contoh tampilan data mentah yang akan menjadi input di R Studio.

Tabel 6. Contoh Tampilan Data Mentah

Jurusan	Asal Sekolah	Jenis	Kelompok	Kota	Provinsi
Akuntansi	SMA HANG TUAH 2	SMA	1	SIDOARJO	JAWA TIMUR
Akuntansi	SMA SENOPATI	SMA	1	SIDOARJO	JAWA TIMUR
Akuntansi	SMAS ST MARKUS RUTENG	SMA	1	MANGGARAI	NUSA TENGGARA TIMUR
Manajemen	SMAN 2 BANJARBARU	SMA	1	BANJAR BARU	KALIMANTAN SELATAN
Manajemen	SMAK KARITAS 3	SMA	1	SURABAYA	JAWA TIMUR
Manajemen	SMA MUHAMMADIYAH 3	SMA	1	SURABAYA	JAWA TIMUR
Hukum	SMA ST. BONAVETURA	SMA	1	MADIUN	JAWA TIMUR
Hukum	SMA KR. PETRA 2	SMA	1	SURABAYA	JAWA TIMUR
Hukum	SMAN 1 BLITAR	SMA	1	BLITAR	JAWA TIMUR
Arsitektur	SMA SANTA MARIA 3	SMA	1	CIMAHI	JAWA BARAT
Arsitektur	SMI	SMI	3	SURABAYA	JAWA TIMUR
Industri	SMA BUNDA HATI KUDUS	SMA	1	SURABAYA	JAWA TIMUR
Industri	SMK ST. LOUIS	SMK	2	SURABAYA	JAWA TIMUR

Pada Tabel 6 diatas dapat dilihat bahwa pengelompokkan sudah dilakukan pada jenis sekolah asal, dimana skor 1 adalah SMA, skor 2 adalah SMK, dan skor 3 adalah lainnya. Kemudian untuk pengelompokkan provinsi penulis telah merekap dari keseluruhan data yang berjumlah 1064 data, nantinya provinsi itu yang akan digunakan, untuk provinsi sumatra utara, sumatera selatan akan digabungkan menjadi provinsi Sumatera, karena data pendaftar yang berasal dari provinsi tersebut hanya 1-3 pendaftar saja, sama halnya dengan kalimantan utara, kalimantan barat, akan digabungkan menjadi provinsi kalimantan saja, dan seterusnya, adapun rekap seluruh provinsi dari 1064 data (data 2015 – 2022) dapat dilihat pada Tabel 7 berikut.

Tabel 7. Rekap Data Provinsi Yang Digunakan



Jurusan	Akuntansi	Manajemen	Hukum	Industri	Arsitektur
Bali	2	2	0	1	1
Jawa Barat	2	1	1	0	1
Jakarta	0	1	0	0	0
Ntt	24	30	16	3	5
Sumatra	5	10	10	3	3
Kaltim	1	3	1	0	3
Jawa Tengah	5	8	0	1	0
Kalimantan	0	2	2	0	0
Papua	0	5	3	0	0
Sulawesi	6	6	1	3	0
Timor Leste	0	0	0	0	1
Jawa Timur	282	357	111	86	55
Maluku	0	1	0	0	0
Total	327	426	145	97	69
Grand Total			1064		

Setelah merapikan data pada excel, maka langkah selanjutnya adalah masuk ke software R Studio untuk memulai melakukan tahap running. Untuk penulisan syntax pada lembar kerja R Studio dapat dilihat dibawah ini.

```
data <- rekap
data
View(rekap) #melihat data
summary(rekap) #ringkasan data
attach(rekap) #memisahkan variabel

index<-data
kluster <- index[2:6]

a <- scale(kluster)
jarak <- get_dist(a)
fviz_nbclust(kluster, kmeans, method = "wss")
fviz_nbclust(kluster, kmeans, method = "silhouette")
fviz_dist(jarak, gradient=list(low="#00AFBB",
mid="white", high="#FC4E07"))
k3 <- kmeans(a, centers = 2, nstart = 10) #penentuan jumlah cluster
k3
fviz_cluster(k3, data=kluster) #visualisasi hasil pengelompokan

table(k3$kluster, rekap$jurusan)
final=data.frame(rekap, kluster$kluster)
final
```

Interpretasi Output Hasil

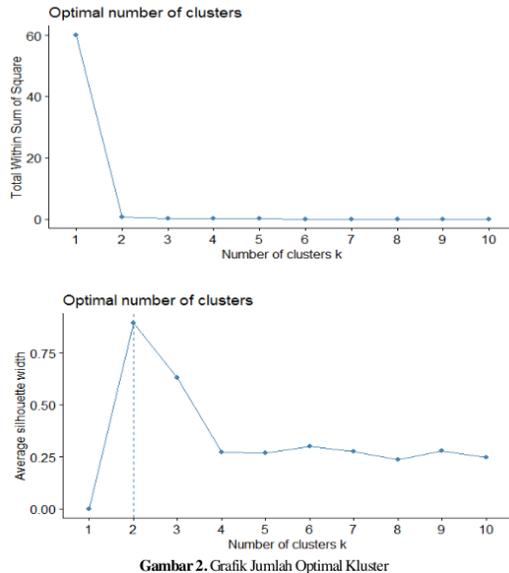
Setelah seluruh syntax yang ada dimasukkan/ dituliskan pada lembar kerja R Studio, maka langkah selanjutnya adalah melakukan interpretasi dari hasil running yang dihasilkan. Terdapat 3 output berupa gambar/visualisasi dari data yang telah di running. Output yang pertama dengan syntax:

```
fviz_nbclust(kluster, kmeans, method = "wss")
fviz_nbclust(kluster, kmeans, method = "silhouette")
```

Menghasilkan Gambar 2 seperti dibawah ini, dimana output ini merupakan grafik yang akan menunjukkan berapa jumlah optimal kluster yang terbentuk dari persebaran data pendaftaran yang diolah. Gambar grafik yang pertama menggunakan metode wss, dimana dengan metode ini terlihat grafik mulai turun drastis pada angka 2, hal ini berarti bahwa



jumlah kluster yang optimal dibentuk adalah 2 kluster saja. kemudian untuk gambar grafik yang kedua adalah menggunakan metode silhouette, hasil yang ditunjukkan menggunakan metode ini adalah 2 kluster, terlihat ada garis putus-putus yang berada pada angka 2, yang berarti bahwa kluster optimal yang terbentuk adalah 2 saja. Secara umum dapat dikatakan bahwa persebaran data pendaftaran mahasiswa memiliki 2 kelompok data yang berbeda, dimana setiap kelompok merupakan data yang memiliki kesamaan dalam hal persebaran datanya.



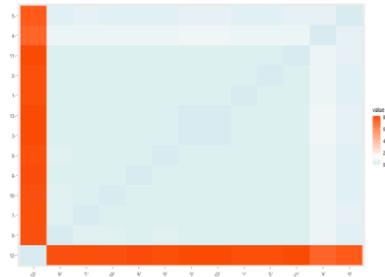
Gambar 2. Grafik Jumlah Optimal Kluster

Output selanjutnya adalah grafik Heat Mape yang dapat dimunculkan dengan menuliskan syntax seperti dibawah ini.

```
fviz_dist(jarak, gradient=list(low="#00AFBB",
mid="white", high="#FC4E07"))
```

Grafik *Heat Mape* merupakan output yang menampilkan visualisasi persebaran data dengan menggunakan warna. Warna merah menunjukkan persebaran datanya semakin jauh, sedangkan jika warnanya semakin hijau maka persebaran datanya semakin dekat. Untuk mengetahui data apa saja yang saling dekat dan jauh dapat dilihat pada garis kordinat X dan Y nya, dimana terdapat angka-angka yang menempel pada garis, angka tersebut merupakan informasi data ke... (contoh: data ke 7, data ke 13, dan sebagainya). Adapun tampilan grafik *heat mape* dapat dilihat pada Gambar 3 dibawah ini.



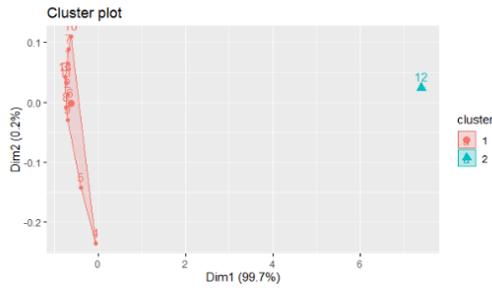


Gambar 3. Grafik Heat Map

Syntax dibawah ini akan menampilkan visualisasi pemetaan persebaran data pendaftar UKDC yang dibedakan dari beberapa provinsi dan jurusan, adapun syntax yang digunakan adalah sebagai berikut.

```
k3 <- kmeans(a, centers = 2, nstart = 10) #penentuan jumlah cluster
k3
fviz_cluster(k3, data=kluster) #visualisasi hasil pengelompokan
```

Mengikuti saran yang muncul pada Gambar 3 diatas, bahwa jumlah cluster yang optimal pada data ini adalah 2, maka penulis juga menuliskan 2 cluster pada syntax. Dari hasil yang diperoleh dapat dilihat bahwa cluster 1 diwakili oleh warna merah, kemudian cluster yang kedua diwakili oleh warna biru. Terdapat 1 data saja yang masuk pada cluster 2 yakni data nomor 12, dan data yang lainnya masuk pada cluster 1. Hal ini berarti persebaran data pada data nomor 12 terlalu jauh jika dibandingkan dengan data lainnya. Untuk lebih jelasnya dapat dilihat pada Gambar 5 dibawah ini, kemudian untuk gambar plot cluster dapat dilihat pada Gambar 4 dibawah ini.



Gambar 4. Plot Kluster

Syntax yang terakhir menunjukkan hasil dalam tabel, untuk lebih memudahkan pembaca dalam melakukan interpretasi data, adapun syntax yang digunakan adalah seperti dibawah ini, kemudian untuk tabel hasil running ada pada gambar 5.

```
table(k3$kluster, rekap$jurusan)
final=data.frame(rekap, kluster=$cluster)
final
```



```

> final=data.frame(rekap, kLaster$cLuster)
> final
  jurusan Akuntansi Manajemen Hukum Industri Arsitektur kLaster.cLuster
1      Bali          2          1          0          1          1          1
2 Jawa Barat          2          1          1          0          1          1
3      Jakarta          0          1          0          0          0          1
4          NTT          24          30          16          3          5          1
5 Sumatra Utara          5          10          10          3          3          1
6      Kaltim          1          3          1          0          3          1
7 Jawa Tengah          5          8          0          1          0          1
8      Kalbar          0          2          2          0          0          1
9          Papua          0          5          3          0          0          1
10 Sulawesi          6          6          1          3          0          1
11 Timor Leste          0          0          0          0          1          1
12 Jawa Timur          282          357          111          86          55          2
13      Maluku          0          1          0          0          0          1

```

Gambar 5. Tabel Hasil Running

Pada gambar diatas terlihat bahwa kluster 1 dianggotai oleh 12 data, dengan pendaftar tertinggi selama tahun 2015-2022 berasal dari Nusa Tenggara Timur dengan jumlah mahasiswa akuntansi 24 orang, manajemen 30 orang, hukum 16 orang, teknik industri 3 orang, dan arsitektur 5 orang. 12 provinsi ini memiliki kesamaan persebaran data maka dari itu seluruhnya masuk ke dalam kluster 1. Sedangkan untuk kluster 2 hanya di isi oleh data nomor 12 yakni provinsi jawa timur. Seperti yang dapat dilihat jumlah pendaftar yang berasal dari jawa timur sangat tinggi, sehingga terlihat jauh sekali perbedaannya dengan provinsi lain, tidak heran jika provinsi jawa timur menjadi kluster tertinggi atau kluster sendiri. secara umum jika dilihat dari data, pendaftar yang berasal dari jawa timur didominasi oleh kota Surabaya dan Sidoarjo.

Simpulan

Setelah melakukan analisa diatas maka penulis dapat menarik kesimpulan bahwa total 1064 data pendaftaran mahasiswa baru sejak tahun 2015-2022 yang dianalisis menggunakan metode *K-Means Clustering* menghasilkan 2 kluster yang optimal, dimana setiap kluster pasti memiliki persebaran data yang hampir mirip maka dari itu mereka dikelompokkan menjadi 1 kluster yang sama. Kluster 1 dianggotai oleh 12 data (provinsi), dengan provinsi pendaftar tertinggi berasal dari Nusa Tenggara Timur dengan total 78 mahasiswa, provinsi Nusa Tenggara ini mencakup Kabupaten Flores, Kabupaten Ende dan Kabupaten Manggarai. Sementara untuk kluster 2 merupakan cluster yang memiliki jumlah data paling banyak dibandingkan dengan kluster 1, dan terlihat adanya selisih perbedaan data yang cukup jauh antara kluster 1 dengan kluster 2. Kluster 2 hanya memiliki 1 anggota saja yakni provinsi Jawa Timur (Kota Surabaya, Kabupaten Sidoarjo, dan sekitarnya), dengan total data pendaftar sebanyak 891 mahasiswa, dimana dalam provinsi ini Kota Surabaya menjadi kota yang paling banyak menyumbang mahasiswa. Dengan adanya hasil ini diharapkan unit PMB dapat lebih menekan upaya promosi kampus ke Provinsi Jawa Timur khususnya Kota Surabaya, dan Provinsi Nusa Tenggara Timur, misalnya dengan cara menitipkan brosur dengan sistem relasi kepada mahasiswa untuk dibawa pada saat pulang kampung, atau sedang berkunjung ke sekolah SMA/ SMK sebelumnya.

Untuk memperluas hasil analisis penelitian ini, maka dari itu penulis selanjutnya dapat melakukan penelitian yang bertujuan untuk melakukan analisis kluster yang lebih spesifik yakni dengan melakukan pengelompokkan data berdasarkan daerah (kabupaten/ kota).

Daftar Pustaka

- [1] A. Wardhana, "Strategi Promosi Penerimaan Siswa Baru (Studi Kasus Smk Kesehatan Cipta Bhakti Husada Yogyakarta)," *CHANNEL J. Komun.*, vol. 6, no. 1, p. 96, 2018, doi: 10.12928/channel.v6i1.10215.
- [2] F. I. Nurlita, *Analisis SWOT Sebagai Strategi Meningkatkan Daya Saing Usaha Kerajinan Sapu Glagah*. 2020.
- [3] T. Triyono, D. Dasmadi, and A. F. A. TNK, "Pengaruh Promosi, Biaya, Fasilitas, Akreditasi, Dan Lokasi Universitas Boyolali Terhadap Minat Calon Mahasiswa Baru," *EKOBIS J. Ilmu Manaj. dan Akunt.*, vol. 9, no. 2, pp. 220–229, 2021.
- [4] Y. M. LAROSA, "Analisis Faktor-Faktor Yang Mempengaruhi Motivasi Mahasiswa Dalam Memilih Konsentrasi Jurusan Manajemen Di Stie Pembnas Nias," *J. Akunt. dan Manaj. Pembnas*, vol. 8, no. 1, pp. 53–62, 2021.
- [5] E. D. S. Mulyani, C. R. Hidayat, and T. C. Ulfa, "Sistem Pakar Untuk Menentukan Jurusan Kuliah Berdasarkan Minat dan Bakat Siswa SMA Dengan Menggunakan Metode Forward Chaining," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 10, no. 2, p. 80, 2021, doi: 10.22303/csr.10.2.2018.80-92.



