

Penerapan *K-Means Clustering* Pada Data Mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata Untuk Menentukan Strategi Promosi

Rioldy Leonard Pattipeilohy^{1*}, Magdalena A. Ineke Pakereng²

^{1,2}Jurusan Teknik Informatika, FTI UKSW, Salatiga, Indonesia

e-mail : 1672018377@student.uksw.edu , ²ineke.pakereng@uksw.edu

Abstract

The research was conducted to see the trend of supply of students studying at the Interdisciplinary Faculty, Tourism Destination D4 Study Program. The algorithm used in the student supply process is K-means. Data processing with the K-means algorithm helps extract information and knowledge from student data in the Interdisciplinary Faculty, D4 Tourism Destination Study Program. By using data mining, stakeholders in the Interdisciplinary Faculty, D4 Tourism Destination Study Program, can take strategic steps in the screening process in provinces that are indicated to supply students. The K-means algorithm facilitates the process of data analysis and grouping of student data for 4 years. The purpose of this research is to provide an accurate and strategic picture of the provinces that can have a significant impact on the supply of students each year. The results showed that the largest supply of students came from Central Java and the smallest from Bangka Belitung, West Sulawesi, Banten, East Kalimantan, West Papua, Bengkulu and Riau, so promotion strategies need to be improved in areas with the smallest student supply.

Keywords: Clustering, K-means, Data Mining

Abstrak

Penelitian yang dilakukan untuk melihat trend suplai mahasiswa yang kuliah di Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata. Algoritma yang digunakan dalam proses suplai mahasiswa ialah K-means. Pengolahan data dengan algoritma K-means ini membantu mengekstrak informasi dan pengetahuan dari data mahasiswa yang ada di Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata. Dengan menggunakan data mining, pemangku kepentingan di Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata, dapat mengambil langkah strategis dalam proses penjangkauan di provinsi-provinsi yang terindikasi mensuplai mahasiswa. Algoritma K-means memfasilitasi proses analisis data dan pengelompokan data mahasiswa selama 4 tahun. Tujuan dari penelitian ini adalah untuk memberikan gambaran yang tepat sasaran dan strategis mengenai provinsi-provinsi yang dapat memberikan dampak signifikan terhadap suplai mahasiswa setiap tahunnya. Hasil penelitian menunjukkan bahwa suplai mahasiswa terbanyak berasal dari daerah Jawa Tengah dan terkecil dari Bangka Belitung, Sulawesi Barat, Banten, Kalimantan Timur, Papua Barat, Bengkulu dan Riau, sehingga strategi promosi perlu ditingkatkan pada daerah-daerah tersebut.

Kata Kunci: K-Means Clustering, Data mining

1. PENDAHULUAN

Universitas Kristen Satya Wacana (UKSW) adalah universitas swasta yang terletak di wilayah Salatiga, Jawa Tengah. Mahasiswa UKSW berasal dari berbagai provinsi di Indonesia, dari Sabang sampai Merauke, karena hal tersebut UKSW Wacana disebut Indonesia Mini. Mahasiswa baru yang terdaftar setiap tahunnya, menghasilkan data dalam bentuk informasi profil baru. Data dan informasi yang



dapat diperoleh dengan mengolah data berdasarkan data mahasiswa. Untuk memperoleh informasi penting berupa pengetahuan baru, maka perlu dilakukan pengolahan data mahasiswa. Informasi yang akan diolah dapat digunakan guna penunjang dalam pengambilan keputusan bagi institusi.

Memproses data dalam jumlah besar, termasuk kumpulan data yang banyak dan data dalam jumlah besar, tidaklah mudah. Salah satu cara untuk mengelola data dengan jumlah record yang besar adalah dengan mengimplementasikan data mining. Teknologi data mining merupakan salah satu alat untuk mengekstrak data dari database besar dengan spesifikasi yang sangat kompleks. Data mining sendiri memiliki metode dan teknik untuk memenuhi kebutuhan, termasuk kebutuhan akan informasi yang luas, dan informasi yang dihasilkan dapat digunakan sebagai alat pengambilan keputusan atau digunakan untuk meningkatkan kualitas proses pengambilan keputusan. Data atau informasi memiliki potensi yang besar untuk dijadikan sebagai kesimpulan pengambilan keputusan dengan menganalisis data yang dikandungnya.

Teknik *data mining* yang akan digunakan dalam penelitian adalah *Clustering*. *Clustering* merupakan salah satu teknik *data mining* yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam *cluster (group)* sehingga setiap *cluster* akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya [1].

Tujuan dari pengolahan data mahasiswa program studi D4 Destinasi Pariwisata Sarjana Interdisiplin adalah untuk menerapkan algoritma clustering *K-Means* pada data penerimaan mahasiswa baru. Atribut data yang digunakan adalah nama siswa, dan wilayah. Manfaat yang diharapkan adalah departemen promosi UKSW akan membantu pelaksanaan promosi serta mencari calon mahasiswa baru.

Algoritma yang digunakan untuk menentukan strategi pemasaran dalam penelitian ini adalah algoritma *K-Means*. Algoritma *K-Means* menjadi salah satu algoritma dalam bidang data mining karena kelebihanannya yaitu implementasi algoritma yang mudah, waktu komputasi yang relatif cepat, dan banyak digunakan untuk menyelesaikan berbagai permasalahan algoritma. Memecahkan Masalah Komputasi Ketika metode pengelompokan dan *K-means* yang paling umum digunakan, data harus ditransfer agar dapat diproses. Setelah itu, perlu menginisialisasi data tipe data nominal, seperti data siswa, dalam bentuk numerik. Jika data sudah dalam bentuk numerik, maka akan diproses menggunakan alat *Orange* [2].

2. METODOLOGI PENELITIAN

2.1. Tinjauan Pustaka

Penelitian yang berjudul *Clustering* Data Mahasiswa Menggunakan Algoritma *K-Means* untuk Menunjang Strategi Promosi, membahas tentang memproses data dalam jumlah besar dengan banyak kumpulan data yang banyak atribut tidaklah mudah. Salah satu cara untuk mengolah data dalam jumlah besar adalah dengan menggunakan *data mining* [1].

Penelitian yang berjudul Penentuan Strategi Promosi Universitas Budi Darma dengan Algoritma *K-Means Clustering*, membahas tentang masalah keputusan



strategi promosi didasarkan pada jumlah siswa dengan latar belakang regional atau universitas. Hal ini menjadi acuan bagi para pengambil keputusan untuk menentukan strategi promosi mereka di Universitas Buddhaarma. Teknik *data mining* dapat diterapkan ketika masalah penentuan strategi periklanan dapat diselesaikan. *Data mining* itu sendiri adalah proses menemukan pola atau informasi baru dalam kumpulan data organisasi atau perusahaan dan menggunakan pola atau informasi baru tersebut sebagai acuan atau dasar pengambilan keputusan [3].

Penelitian yang berjudul Penerapan Algoritma *K-Means Clustering* Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Hasyim Asy'Ari Jombang, dan File Dokumen menggunakan algoritma *K-Means*, membahas tentang membuat analisa data menggunakan metode *clustering* dalam data mining menggunakan algoritma *K-Means*. Algoritma *K-Means* sudah sangat banyak digunakan oleh para peneliti sebelumnya untuk mendapatkan berbagai informasi yang digunakan untuk menentukan strategi periklanan dengan mengelompokkan data [4].

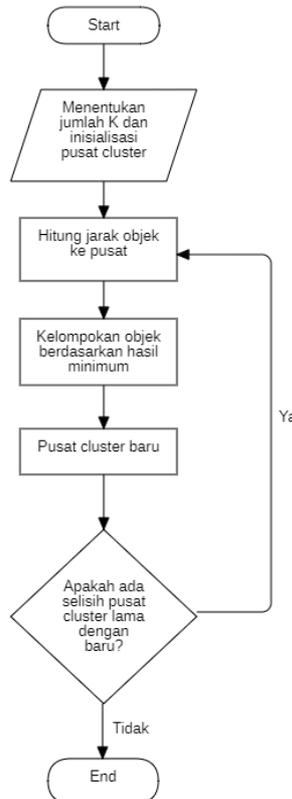
Penelitian yang berjudul Algoritma *K-Means Clustering* Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia, membahas tentang masalah data mahasiswa. Data tersebut perlu dikumpulkan dan dikelompokkan berdasarkan alamat asal sekolah dan jurusan. Pengelompokan data dengan menggunakan *K-Means Clustering*. Algoritma yang digunakan yaitu *K-Means* agar data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [5].

Penelitian yang berjudul Analisis Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing dengan studi kasus Hoyweapstore. Data tersebut melakukan penerapan algoritma *K-Means* menggunakan aplikasi *Orange studio* versi 8.2.001 ini berguna saat melakukan aktivitas seperti *data mining*, pengelompokan, klasifikasi, dan regresi. [2].

Clustering adalah salah satu teknik data mining yang dikenal. Pengertian *clustering* dalam *data mining* adalah mengelompokkan sekumpulan data atau objek ke dalam klaster (kelompok) sehingga semua data dalam suatu klaster berisi data yang semirip mungkin dengan objek di klaster lain saja. Dua metode pengelompokan dikenal: pengelompokan hierarkis dan partisi. Metode hirarki *clustering* sendiri terdiri dari *complete linkage clustering*, *single linkage clustering*, *mean linkage clustering*, dan *centroid linkage clustering*. Metode segmentasi sendiri terdiri dari *k-means* dan *fuzzy k-means*. *Clustering* adalah bagian dari ilmu *data mining* tanpa pengawasan. *Clustering* adalah proses membagi data ke dalam kelas atau *cluster* berdasarkan kesamaan.

Untuk pengelompokan *K-means*, K dimaksudkan sebagai konstanta untuk jumlah *cluster* yang diinginkan. *k-means clustering* merupakan metode analisis data atau metode *data mining* karena *mean* dalam hal ini berarti *mean* dari sekelompok data, yang dalam hal ini didefinisikan sebagai *cluster*. Suatu metode menjalankan proses pemodelan tanpa pengawasan (*unsupervised*) dan menggunakan sistem partisi untuk melakukan pengelompokan data. [1]. Alur proses Algoritma *K-Means Clustering* dinyatakan dalam bentuk *flowchart*, yang

dapat dilihat pada Gambar 1. Suatu metode menjalankan proses pemodelan tanpa pengawasan (*unsupervised*) dan menggunakan sistem partisi untuk melakukan pengelompokan data. [1]. Alur proses *Algoritma K-Means Clustering* dinyatakan dalam bentuk *flowchart*, yang dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* Alur Proses Algoritma K-Means

Alur proses algoritma K-means pada Gambar 1, dijelaskan sebagai berikut :

- a) Menentukan jumlah *cluster* yang akan dibentuk
- b) Inisialisasi k pusat *cluster* dan memberi nilai awal dengan angka-angka acak

$$D(i, j) = \sqrt{(X_{1i} - Y_{1j})^2 + (X_{2i} - Y_{2j})^2 + \dots + (X_{ki} - Y_{kj})^2} \quad (1)$$

Keterangan:

$D(i, j)$ = Jarak objek i ke pusat *cluster* j

X_{ki} = Data ke i variabel data ke k

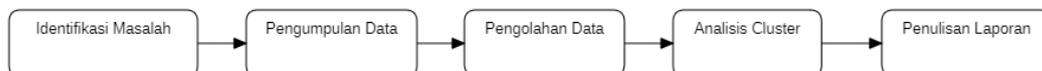
X_{kj} = Titik pusat ke j pada variabel ke k

- c) Kelompokkan data ke dalam *cluster* berdasarkan jarak yang paling dekat.
- d) Mencari pusat *cluster* baru dengan cara menghitung rata-rata dari data objek dalam *cluster* yang terbentuk
- e) Kembali ke langkah c sampai dengan e sampai tidak ada lagi objek yang berpindah ke *cluster*.

2.2. Metode Penelitian

Data mining adalah analisis pemeriksaan kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang sebelumnya tidak diketahui, sehingga dapat dimengerti dan berguna bagi pemilik data. *Data mining* adalah bidang interdisiplin yang menggabungkan pembelajaran mesin, pengenalan pola, statistik, *database*, dan teknik visualisasi untuk memecahkan masalah pengambilan informasi dari *database* besar. *Data mining* adalah istilah yang digunakan untuk menggambarkan pencarian pengetahuan dalam *database*. *Data mining* adalah proses mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan gabungan dari *database* besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Berdasarkan pengertian data mining sebelumnya, *data mining* adalah pengetahuan yang tersembunyi dalam *database*, yang menggunakan matematika, kecerdasan buatan, dan pembelajaran mesin untuk menemukan pola dan teknik statistik untuk mengekstrak dan mengidentifikasi informasi pengetahuan dari database [3].

Metode Penelitian yang digunakan dalam penerapan algoritma *K-Means* untuk menentukan strategi promosi pada data mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata menggunakan rancangan penelitian yang ditunjukkan pada Gambar 2 yaitu: Identifikasi Masalah (1), Pengumpulan Data(2), Pra-Proses Data (3), Analisis *Cluster* (4), dan Penulisan Laporan (5).



Gambar 2 Rancangan Penelitian

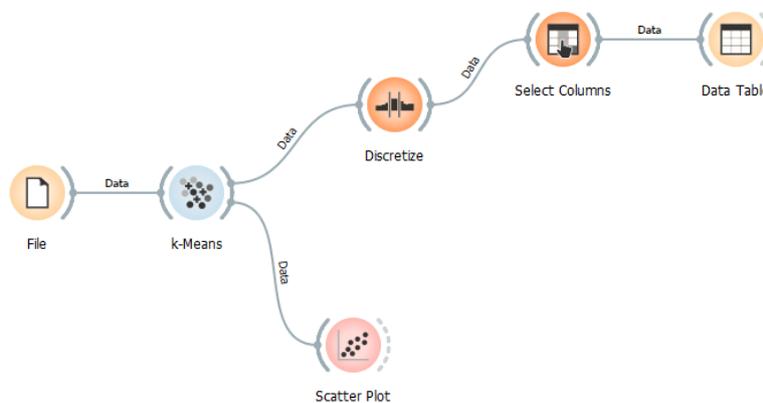
Rancangan penelitian pada Gambar 2, dijelaskan sebagai berikut. Tahap Identifikasi Masalah, Identifikasi Masalah adalah salah satu langkah penting dalam melakukan penelitian untuk menemukan jawaban dari suatu masalah untuk mengetahui daerah yang harus mendapatkan promosi; Tahap Pengumpulan Data, Data yang akan digunakan dalam penelitian akhir ini adalah data mahasiswa yang masuk Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata. Data yang akan diseleksi berisi tentang data mahasiswa seperti nama, asal sekolah, asal daerah. Dalam penelitian ini digunakan beberapa metode penelitian yaitu: Wawancara merupakan metode pengumpulan data dengan melakukan tanya jawab secara langsung dengan civitas akademika Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata, dan Studi literatur adalah mempelajari dan memahami karya ilmiah ataupun sumber ilmiah yang lain yang sesuai dengan penelitian dan memiliki hubungan penelitian yang serupa [2]; Tahap Pengolahan Data, yaitu mengolah data yang diberikan Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata, lalu data tersebut diolah dengan metode *K-Means Clustering*; Tahap Analisis Kluster, *K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam



satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Penerapan proses algoritma *K-Means* menggunakan aplikasi *Orange studio* yang berguna dalam membantu melakukan kegiatan seperti data mining, klustering, klasifikasi, dan regresi; dan Tahap Penulisan Laporan, pada tahap ini hasil penelitian yang selesai dilakukan kemudian ditulis ke dalam artikel ilmiah atau laporan penelitian.

3. HASIL DAN PEMBAHASAN

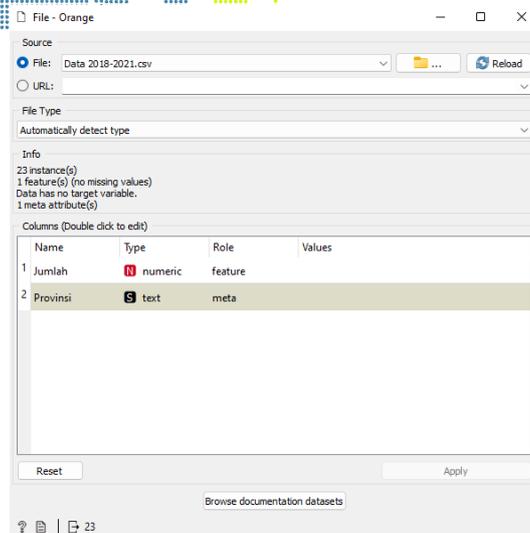
Pada proses pengolahan data digunakan metode *clustering* dengan algoritma *K-means*. Proses pengolahan dilakukan menggunakan 4 (empat) *cluster* untuk mengelompokkan data selama kurun waktu 4 (empat) tahun. Penggunaan 4 (empat) buah *cluster* mempermudah untuk menentukan skala prioritas, potensi suplai, serta batasan atau ruang lingkup pada data 4 (empat) tahun dalam waktu 2018-2021.



Gambar 3. Proses Pengolahan Program *Orange*

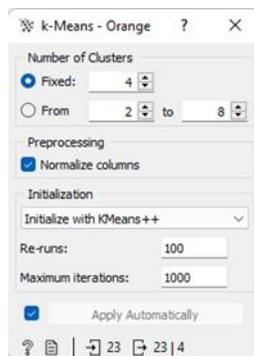
Gambar 3 merupakan proses pengolahan *data mining* menggunakan aplikasi *Orange*. Ada beberapa *tools* yang digunakan dalam proses pengolahan pada gambar 3, antara lain:

- a) *File* : di dalam *tools file* terdapat data dengan ekstensi *.csv* yang telah dikelompokkan sebelumnya.
- b) *K-Means* : merupakan algoritma yang digunakan untuk mengolah data. Dalam proses *K-Means*, ditentukan untuk jumlah *cluster*.
- c) *Scatter Plot* : *tool* yang digunakan untuk visualisasi hasil pengelompokan.
- d) *Discretize* : *tool* untuk memberikan batasan nilai pada data
- e) *Select Columns* : digunakan untuk mengolah hasil *attribute* serta *features* dari *file* data.
- f) *Data Table* : digunakan untuk melihat hasil dari proses *cluster* beserta dengan *silhouette*.



Gambar 4. Data Mahasiswa Tahun 2018-2021

Gambar 4 menunjukkan data .csv selama periode 4 (empat) tahun, yang telah dinormalisasi untuk menghapus data *outlier* dan data yang tidak memiliki nilai. Dua atribut ditampilkan, pertama adalah jumlah yang terbaca sebagai *numeric*, serta data provinsi terbaca sebagai *text*.



Gambar 8. Proses Untuk Menentukan *Cluster*

Gambar 5 menunjukkan proses penentuan jumlah *cluster*. Pada Gambar 5, proses pengolahan data menggunakan 4 (empat) *cluster* dan dengan proses pengolahan data dengan proses maksimum iterasi sebanyak 1000 kali. Hasil pengolahan data 4 (empat) tahun, memiliki atribut prefektur atau acuan yaitu jumlah Mahasiswa, dan tahun pengolahan dengan menginisialisasi 4 buah *cluster*. Hasil *cluster* juga menunjukkan provinsi lain memiliki potensi yang cukup besar dalam hal penyediaan mahasiswa pada tahun-tahun yang mendatang

Tabel 1 Hasil *Clustering* Data 2018-2021

Provinsi	Cluster	Jumlah	Silhouette
Jawa Tengah	C2	44	0.500
Papua	C3	23	0.500
Lampung	C4	15	0.590



Provinsi	Cluster	Jumlah	Silhouette
Jawa Barat	C4	11	0.684
Nusa Tenggara Timur	C4	9	0.660
Maluku	C4	8	0.609
Kalimantan Tengah	C1	5	0.647
Sulawesi Utara	C1	4	0.697
Maluku Utara	C1	4	0.697
Kalimantan Barat	C1	4	0.697
Jawa Timur	C1	3	0.719
Sulawesi Tengah	C1	3	0.719
Sumatera Selatan	C1	3	0.719
Sumatera Utara	C1	2	0.726
Sumatera Barat	C1	2	0.726
Kalimantan Utara	C1	2	0.726
Bangka Belitung	C1	1	0.725
Sulawesi Barat	C1	1	0.725
Banten	C1	1	0.725
Kalimantan Timur	C1	1	0.725
Papua Barat	C1	1	0.725
Bengkulu	C1	1	0.725
Riau	C1	1	0.725

Data pada Tabel 1 menunjukkan 23 provinsi dengan pasokan bervariasi selama 4 tahun terakhir. Dengan jumlah tersebut, Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata memiliki batasan dalam menjalankan proses promosi, serta penjangkauan pada semua provinsi tersebut. Dengan menggunakan model pengelompokan *K-means* untuk fokus pada wilayah yang diminati, dan fokus pada proses penjangkauan. Pada Tabel 1, provinsi-provinsi dibagi berdasarkan kemiripan nilai, dan dibagi menjadi 4 *cluster* dengan label C1, C2, C3, dan C4.

Tabel 2. Cluster Provinsi Terbanyak Pertama

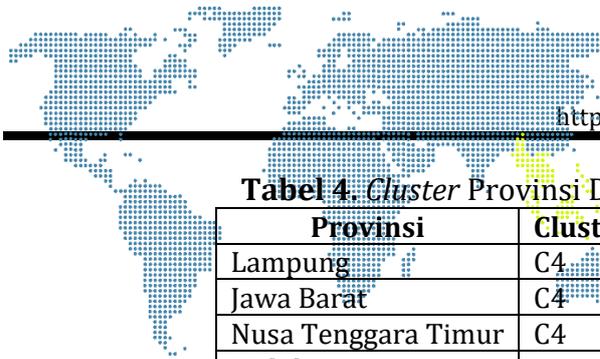
Provinsi	Cluster	Jumlah	Silhouette
Jawa Tengah	C2	44	0.500

Berdasarkan data pada Tabel 2, *cluster* C2 adalah *cluster* dengan nilai terbesar di antara *cluster* lainnya. Pada *cluster* C1, Jawa Tengah memiliki kontribusi mahasiswa terbesar di antara seluruh suplai untuk setiap provinsi yang terdapat pada Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata dalam kurun waktu 4 tahun, dengan jumlah 44 mahasiswa.

Tabel 3. Cluster Provinsi Terbanyak Kedua

Provinsi	Cluster	Jumlah	Silhouette
Papua	C3	23	0.500

Berdasarkan data pada Tabel 3, *cluster* C3 merupakan *cluster* dengan nilai suplai terbesar kedua setelah Jawa Tengah. Hanya terdapat satu provinsi dengan label *cluster* C3 yaitu Papua dengan jumlah 23 mahasiswa dengan nilai *silhouette* adalah 0.500.



Tabel 4. Cluster Provinsi Dengan Nilai Menengah

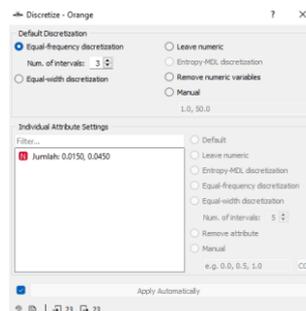
Provinsi	Cluster	Jumlah	Silhouette
Lampung	C4	15	0.590
Jawa Barat	C4	11	0.684
Nusa Tenggara Timur	C4	9	0.660
Maluku	C4	8	0.609

Berdasarkan data pada tabel 4, *cluster* C4 adalah *cluster* dengan nilai menengah. Suplai rata-rata di angka 8 hingga 15 mahasiswa. Pada *cluster* C5 terdapat beberapa provinsi antara lain Lampung, Jawa Barat, Nusa Tenggara Timur, dan Maluku.

Tabel 5. Cluster Provinsi Dengan Nilai Terkecil

Provinsi	Cluster	Jumlah	Silhouette
Kalimantan Tengah	C1	5	0.647
Sulawesi Utara	C1	4	0.697
Maluku Utara	C1	4	0.697
Kalimantan Barat	C1	4	0.697
Jawa Timur	C1	3	0.719
Sulawesi Tengah	C1	3	0.719
Sumatera Selatan	C1	3	0.719
Sumatera Utara	C1	2	0.726
Sumatera Barat	C1	2	0.726
Kalimantan Utara	C1	2	0.726
Bangka Belitung	C1	1	0.725
Sulawesi Barat	C1	1	0.725
Banten	C1	1	0.725
Kalimantan Timur	C1	1	0.725
Papua Barat	C1	1	0.725
Bengkulu	C1	1	0.725
Riau	C1	1	0.725

Data pada Tabel 5 merupakan *cluster* dengan nilai suplai terkecil selama kurun waktu 4 tahun. Hasil *cluster* provinsi-provinsi ini memberikan jumlah paling sedikit dibandingkan dengan *cluster* lainnya. *Cluster* C1 merupakan *cluster* dengan jumlah terbanyak dengan total 17 provinsi.



Gambar 12. Proses Pembagian Data Menjadi 3 Interval

Gambar 6 menunjukkan hasil pengolahan untuk *cluster* yang dibagi menjadi 3 (tiga) interval, yaitu suplai prioritas, potensi suplai, dan suplai terendah, dengan memilah nilai jumlah data berupa nama tiap provinsi, *cluster*, dan nilai. Dengan membagi 3 (tiga) interval, Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata dapat memprioritaskan provinsi yang memberikan suplai terbanyak, berpotensi memberikan suplai, dan daerah yang memberikan suplai terkecil. Pembagian 3 (tiga) interval bermanfaat untuk proses strategi promosi ke provinsi-provinsi tersebut.

Tabel 6. Provinsi Prioritas

Provinsi	Cluster	Jumlah
Jawa Tengah	C2	≥ 0.045
Lampung	C4	≥ 0.045
Papua	C3	≥ 0.045
Jawa Barat	C4	≥ 0.045
Maluku	C4	≥ 0.045
Nusa Tenggara Timur	C4	≥ 0.045
Kalimantan Tengah	C1	≥ 0.045

Berdasarkan data pada Tabel 6, dari hasil pemecahan data menjadi 3 (tiga) interval, terdapat 7 (tujuh) provinsi dengan suplai prioritas yang menunjukkan suplai dengan jumlah terbanyak. Jumlah rata-rata dari tiap daerah tersebut lebih atau sama dengan 0.045. Provinsi prioritas adalah C1 meliputi Kalimantan Tengah, C2 meliputi Jawa Tengah, C3 meliputi Papua, dan C4 meliputi Lampung, Jawa Barat, Maluku, dan Nusa Tenggara Timur.

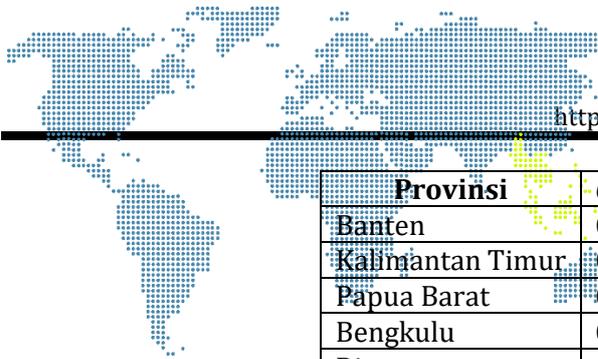
Tabel 7. Provinsi Potensi Suplai

Provinsi	Cluster	Jumlah
Sulawesi Utara	C1	0.015 - 0.045
Jawa Timur	C1	0.015 - 0.045
Maluku Utara	C1	0.015 - 0.045
Sumatera Utara	C1	0.015 - 0.045
Kalimantan Barat	C1	0.015 - 0.045
Sulawesi Tengah	C1	0.015 - 0.045
Sumatera Selatan	C1	0.015 - 0.045
Sumatera Barat	C1	0.015 - 0.045
Kalimantan Utara	C1	0.015 - 0.045

Berdasarkan data pada Tabel 7, provinsi potensi suplai terdiri dari 9 provinsi dengan nilai rata-rata 0.015 - 0.045 yang meliputi *cluster* C1 di antaranya Sulawesi Utara, Jawa Timur, Maluku Utara, Sumatera Utara, Kalimantan Barat, Sulawesi Tengah, Sumatera Selatan, Sumatera Barat, dan Kalimantan Utara.

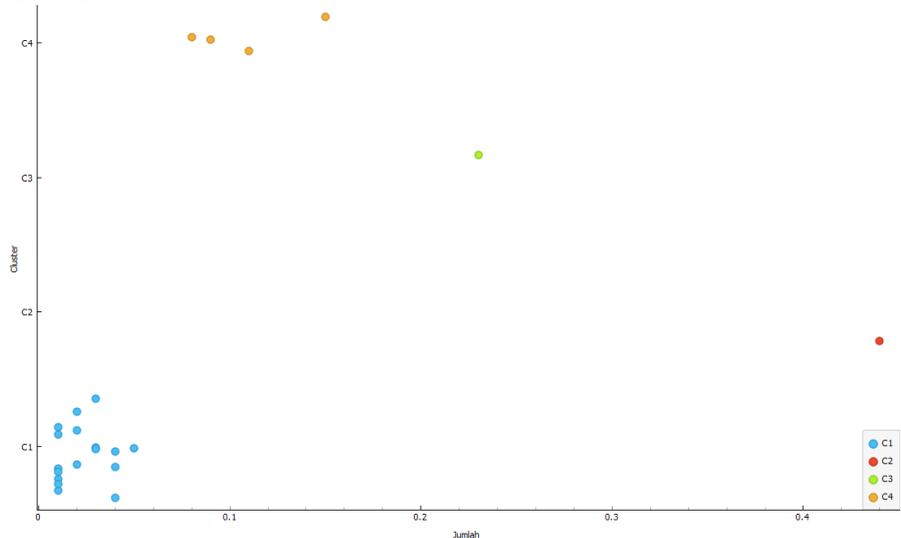
Tabel 8. Provinsi Dengan Suplai Terkecil

Provinsi	Cluster	Jumlah
Bangka Belitung	C1	< 0.015
Sulawesi Barat	C1	< 0.015



Provinsi	Cluster	Jumlah
Banten	C1	< 0.015
Kalimantan Timur	C1	< 0.015
Papua Barat	C1	< 0.015
Bengkulu	C1	< 0.015
Riau	C1	< 0.015

Tabel 8 menunjukkan provinsi dengan suplai terkecil dengan nilai rata-rata lebih kecil dari 0.015. Provinsi di antaranya adalah Bangka Belitung, Sulawesi Barat, Banten, Kalimantan Timur, Papua Barat, Bengkulu, dan Riau yang termasuk dalam *cluster* C1.



Gambar 13. Hasil *Scatter Plot Cluster*

Gambar 7 merupakan hasil visualisasi proses menggunakan empat *cluster*. Terlihat bahwa sebaran tiap prefektur membentuk *cluster* yang saling berdekatan karena adanya kesamaan nilai. Proses tersebut melewati 1000 iterasi, sehingga hasil yang didapatkan tidak mengalami pergeseran posisi, serta jarak dari masing-masing *cluster*.

Penelitian yang dilakukan pada Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata mendapatkan hasil serta memiliki pola menarik dan pengetahuan baru. Studi ini juga menggunakan klasifikasi interval untuk memastikan perhatian yang sama untuk negara bagian dalam setiap klaster yang dibagi menjadi suplai prioritas, potensi prioritas, dan suplai terkecil. Oleh karena itu, dapat diketahui pola untuk setiap keadaan dengan lebih akurat.

Ada strategi yang dapat diterapkan oleh Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata. Pada tiap-tiap provinsi prioritas dan provinsi dengan memiliki nilai potensi suplai memiliki kemungkinan pertambahan suplai mahasiswa. Terdapat 9 provinsi potensi suplai, dan di antaranya hanya satu provinsi yang berada di pulau Jawa. Hal ini sangat berpengaruh kepada tim yang bekerja untuk menjalankan promosi. Karena dari hasil *data mining*, daerah atau provinsi tersebut memberikan potensi suplai paling banyak.



4. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang dilakukan, dapat ditarik kesimpulan bahwa 1) Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata dapat menentukan fokus penyebaran atau penjangkaran dari provinsi dengan potensi suplai terlebih dahulu dan selanjutnya ke provinsi dengan suplai prioritas atau preferensial; 2) dengan menggunakan proses *cluster K-means*, pembagian pada proses mahasiswa pada tiap wilayah provinsi lebih tepat sasaran. Sehingga dengan menggunakan proses *data mining* menggunakan metode *clustering* Fakultas Interdisiplin, Program Studi D4 Destinasi Pariwisata, dapat menerapkan strategi yang efisien serta dapat meminimalkan pengeluaran untuk turun secara eksklusif ke tiap-tiap provinsi; 3) Suplai mahasiswa terbanyak berasal dari daerah Jawa Tengah dan terkecil dari Bangka Belitung, Sulawesi Barat, Banten, Kalimantan Timur, Papua Barat, Bengkulu dan Riau, sehingga strategi promosi perlu ditingkatkan pada daerah-daerah tersebut; dan 4) untuk daerah atau provinsi-provinsi yang belum mendapatkan penjangkaran, diharapkan dari bagian pemasaran atau promosi, dapat melakukan penjangkaran ke tiap-tiap provinsi tersebut. Ada 14 provinsi yang belum atau tidak ada dalam data mahasiswa.

DAFTAR PUSTAKA

- [1] W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari)," *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37.
- [2] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [3] N. Silalahi, "Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means Clustering," *TIN Terap. Inform. Nusant.*, vol. 1, no. 1, pp. 40–46, 2020.
- [4] I. Mahmud, A. D. Indriyanti, and I. Lazulfa, "Penerapan Algoritma K-Means Clustering Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Hasyim Asy'ari Jombang," *Inovate*, vol. 4, no. 2, pp. 20–27, 2020.
- [5] R. Sibarani and O. Omby, "Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia," *J. Algoritma. Log. dan Komputasi*, vol. 1, no. 2, pp. 44–50, 2018, doi: 10.30813/j-alu.v1i2.1367.