

# PENERAPAN *DATA MINING* UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI DI STMIK PRIMAKARA MENGGUNAKAN ALGORITMA *K-MEANS* *CLUSTERING*

Nengah Widya Utami<sup>1)</sup> A.A. Istri Ita Paramitha<sup>2)</sup>

Program Studi Sistem Informasi Akuntansi<sup>1)</sup>

Program Studi Sistem Informasi<sup>2)</sup>

STMIK Primakara, Denpasar, Bali<sup>1) 2)</sup>

widya@primakara.ac.id<sup>1)</sup> ita@primakara.ac.id<sup>2)</sup>

## ABSTRACT

*The process of accepting new students at STMIK Primakara Denpasar produces an abundance of student personal profile data. This will continue every year so that the data stored in the database will also increase in number. It is very unfortunate if the data owned is not used properly. This study applies a data mining method with the k-means clustering algorithm which aims to determine the pattern of study program selection for new students in the STMIK Primakara environment. The data that has been obtained is processed through a pre-processing stage which includes data cleaning, data integration, data selection and data transformation. After going through the pre-processing stage, the next step is the application of data mining using the k-means clustering algorithm. In this stage, data that have the same similarities and characteristics are grouped into certain clusters. The last stage is data interpretation on the pattern of information generated from the mining process. The results showed that after grouping student data based on the study program, regional origin, and type of school using K-Means clustering, 3 clusters were formed, where each cluster has different characteristics based on the study program, type of school, and regional origin. . The results of this study can then be used as a basis for making decisions to determine strategies in promoting study programs at STMIK Primakara.*

**Keywords:** *K-Means, clustering, major, STMIK Primakara.*

## ABSTRAK

Proses penerimaan mahasiswa baru di STMIK Primakara Denpasar menghasilkan data profil pribadi mahasiswa yang jumlahnya melimpah. Hal tersebut akan terus berlangsung setiap tahunnya sehingga data yang tersimpan pada *database* juga akan semakin banyak jumlahnya. Hal tersebut sangat disayangkan apabila data yang dimiliki tidak dimanfaatkan dengan baik. Penelitian ini menerapkan metode *data mining* dengan algoritma *k-means clustering* yang bertujuan untuk mengetahui pola pemilihan program studi bagi mahasiswa baru di lingkungan STMIK Primakara. Data yang telah diperoleh diproses melalui tahap *pre-processing* yang meliputi *data cleaning* (pembersihan data), *data integration* (integrasi data), *data selection* (seleksi data) dan *data transformation* (transformasi data). Setelah melalui tahap *pre-processing* tersebut, langkah selanjutnya adalah penerapan *data mining* menggunakan algoritma *k-means clustering*. Dalam tahap ini, data yang memiliki kemiripan dan karakteristik yang sama dikelompokkan dalam *cluster* tertentu. Tahap terakhir yaitu *data interpretation* (interpretasi data) terhadap pola informasi yang dihasilkan dari proses *mining*. Hasil penelitian menunjukkan bahwa setelah dilakukan pengelompokan data mahasiswa berdasarkan program studi, asal daerah, dan jenis sekolah menggunakan *K-Means clustering* terbentuk 3 (tiga) *cluster*, dimana masing-masing *cluster* memiliki karakteristik yang berbeda berdasarkan program studi, jenis sekolah, dan asal daerah. Hasil dari penelitian ini selanjutnya dapat digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi dalam mempromosikan program studi di STMIK Primakara.

**Kata Kunci :** *K-Means, clustering, program studi, STMIK Primakara.*

## PENDAHULUAN

Proses penerimaan mahasiswa baru dalam sebuah perguruan tinggi tentunya menghasilkan data yang berlimpah berupa profil dari masing-masing mahasiswa seperti nama, tempat dan tanggal lahir, alamat, asal sekolah, jurusan sekolah, nilai, dan yang lainnya. Hal tersebut akan terjadi secara berulang pada sebuah perguruan tinggi setiap tahunnya, sehingga jumlah data pada *database* perguruan tinggi akan semakin meningkat pula. Sama halnya yang terjadi pada perguruan tinggi STMIK Primakara, Denpasar. Penumpukan data mahasiswa secara terus menerus tersebut menyebabkan sulitnya melakukan pencarian informasi terhadap suatu data mahasiswa.

Banyaknya data yang tersimpan dalam *database* STMIK Primakara belum dimanfaatkan dengan baik selain sebagai kebutuhan administrasi. Berdasarkan jumlah data mahasiswa yang melimpah tersebut, sebenarnya terdapat informasi yang tersembunyi yang dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga dapat berguna bagi pihak perguruan tinggi. Data yang terkumpul dari tahun ke tahun bisa di analisis untuk menghasilkan informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen perguruan tinggi dalam membuat strategi promosi yang tepat. Salah satu contohnya adalah pemanfaatan data profil mahasiswa baru yang diperoleh melalui proses pendaftaran atau registrasi yang meliputi nama mahasiswa, jurusan SMA/SMK/MA, kota asal mahasiswa, tempat tinggal, nilai ujian, serta program studi yang dipilih. Selain data mahasiswa yang belum dimanfaatkan dengan baik, permasalahan yang terjadi di STMIK Primakara yakni belum dapat menentukan strategi promosi yang tepat, dimana sasaran/target promosi program studi selama ini menyasar seluruh daerah di Bali dan NTB. Promosi yang dilakukan oleh tim marketing selama ini tidak mengacu pada sebaran peminatan mahasiswa pada setiap program studi di setiap daerahnya. Hal tersebut dianggap kurang efektif dan efisien serta belum memberikan dampak yang signifikan terhadap peningkatan jumlah mahasiswa yang diperoleh. Adapun metode yang dapat digunakan untuk mengolah data profil mahasiswa menjadi

sebuah pengetahuan disebut dengan *data mining*. *Data mining* merupakan langkah analisis terhadap proses penemuan pengetahuan di dalam basis data atau *knowledge discovery in databases* yang disingkat KDD. Pengetahuan yang dimaksud dapat berupa pola data atau relasi antar data yang valid (yang tidak diketahui sebelumnya) [1]. Algoritma yang dapat digunakan dalam pengelompokan data calon mahasiswa baru adalah metode *clustering*. Menurut Asroni, dkk (2018) *clustering* merupakan cara untuk menemukan kelompok objek yang memiliki kemiripan dan dapat menemukan pola penyebaran dan pola hubungan dalam kumpulan data yang besar. Dalam proses *clustering* yang terpenting adalah mengumpulkan pola ke kluster yang sesuai untuk menemukan persamaan dan perbedaan agar menghasilkan kesimpulan yang berharga [2]. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*) yang memiliki arti bahwa karakteristik tiap kluster tidak ditentukan sebelumnya melainkan berdasarkan kemiripan atribut-atribut dari suatu kelompok atau kluster [3].

Salah satu algoritma *clustering* dalam *data mining* adalah algoritma K-Means. Algoritma K-means memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu algoritma *k-means* tidak terpengaruh dengan adanya urutan objek [2]. Dengan menggunakan metode ini, data yang telah didapatkan dapat dikelompokkan kedalam beberapa kluster berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu kluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam kluster yang lain yang memiliki karakteristik yang sama [3].

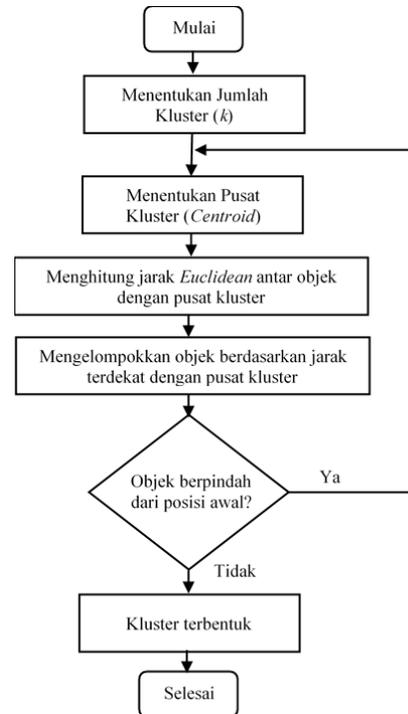
Berdasarkan permasalahan-permasalahan tersebut maka dilakukan penelitian “*Penerapan Data Mining Untuk Mengetahui Pola Pemilihan Program Studi Di STMIK Primakara Menggunakan Algoritma K-Means Clustering*” yang bertujuan untuk mengetahui pola pemilihan program studi di STMIK Primakara, mengetahui pola minat mahasiswa

baru terhadap program studi, dan dapat menjadi rujukan pengetahuan untuk pengelola program studi dan tim marketing dalam memberikan sosialisasi dan promosi ke masyarakat. Dengan demikian strategi promosi program studi menjadi lebih efisien dan tepat sasaran.

**TINJAUAN PUSTAKA**

**K-Means Clustering**

Menurut Rony (2016) *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar [3]. Salah satu algoritma *data mining* adalah *K-Means clustering* yang merupakan metode untuk mengelompokkan data *non-hierarki* yang mempartisi data ke dalam bentuk dua atau lebih kelompok, sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang di *set* dalam suatu kelompok dan memaksimalkan variasi antar kelompok [2]. *K-Means* merupakan algoritma klasterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya [1]. Penerapan algoritma *K-Means* dilakukan dengan menggunakan *tools data mining*. Algoritma *K-Means* merupakan algoritma yang membutuhkan parameter input sebanyak *k* dan membagi sekumpulan *n* objek kedalam *k* kluster sehingga tingkat kemiripan antar anggota dalam suatu kluster tinggi sedangkan tingkat kemiripan dengan anggota pada kluster lain sangat rendah. Kemiripan anggota terhadap kluster diukur dengan kedekatan objek terhadap nilai *mean* pada kluster atau dapat disebut sebagai *centroid* kluster. Pada algoritma *K-Means*, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang diterima adalah data atau objek dan *k* buah kelompok (kluster) yang diinginkan [4]. Tahapan algoritma *K-Means* dapat dilihat pada gambar berikut.



Gambar 1. Tahapan Algoritma K-Means

Berdasarkan gambar di atas, tahapan algoritma *K-Means* meliputi [5]:

- a. Menentukan jumlah *cluster k*
- b. Inisialisasi *k* pusat *cluster (centroid)* dengan cara *random*
- c. Mengalokasikan semua data/objek ke kluster paling dekat. Untuk melakukan proses pengolahan data pada tiap titik pusat *cluster* menggunakan jarak *Euclidean* yang dirumuskan sebagai berikut:
 
$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

$D_{(i,j)}$  = Jarak data ke *i* ke pusat kluster *j*  
 $X_{ki}$  = Data ke *i* pada atribut data ke *k*  
 $X_{kj}$  = Titik pusat ke *j* pada atribut ke *k*
- d. Proses selanjutnya adalah menghitung ulang pusat kluster dengan keanggotaan kluster yang terbaru. rata-rata dari semua data/objek dalam kluster merupakan pusat kluster
- e. Menugaskan kembali setiap objek dengan menggunakan pusat kluster baru, apakah

pusat kluster berubah hingga proses *clustering* selesai, ulang kembali proses “c” sampai menemukan nilai pusat kluster tidak ada yang berubah.

### **Penelitian Terkait**

Penerapan algoritma Apriori untuk mengetahui Beberapa peneliti telah memanfaatkan algoritma K-Means, seperti halnya yang telah dilakukan oleh Khoiriyatus, Yuliansyah, Arfiani (2019) yang berjudul “*Clustering Student Data Based on K-Means Algorithms*”. Penelitian tersebut menggunakan 724 data mahasiswa dan 4 (empat) atribut, yaitu Indeks Prestasi Kumulatif (IPK), lama studi (LS), skor kecakapan bahasa Inggris (EP), dan lama tugas tesis (LT). Hasil dari penelitian ini menghasilkan 3 (tiga) kluster yaitu kluster karakteristik siswa yang berprestasi, berprestasi standar, dan berprestasi rendah [6]. Penerapan algoritma K-Means dalam menemukan pengetahuan pada data akademik perguruan tinggi juga dilakukan oleh Omolewa, Oladele, Adeyinka, Oluwaseun (2019) yang berjudul “*Prediction of Student’s Academic Performance using K-Means Clustering and Multiple Linear Regressions*”. Pada penelitian ini dikembangkan model untuk memprediksi kinerja siswa dengan menggunakan regresi linier berganda dan mengelompokkan mahasiswa menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa nilai ulangan, kuis dan tugas merupakan faktor utama yang dapat digunakan dalam memprediksi prestasi akademik mahasiswa serta diperoleh 2 (dua) kluster mahasiswa berdasarkan pengelompokan prestasi akademik [7].

Selain itu, penelitian lain juga dilakukan oleh Munawaroh, Handayani, dan Wedaningsih (2019) yang berjudul “*The Implementation of K-Means Algorithm for Cluster Majoring to New Students in SMKN 2 of South Tangerang*”. Penelitian tersebut bertujuan untuk membantu siswa dalam menentukan jurusan dan membantu sekolah dalam pengelompokan jurusan. Dalam proses *clustering* dengan algoritma K-Means tersebut menghasilkan informasi jumlah siswa pada setiap jurusan [8]. Penerapan algoritma K-Means untuk mengetahui pola pemilihan jurusan juga

dilakukan oleh Yunita (2018) yang berjudul “*Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru*”. Penelitian tersebut bertujuan untuk melakukan pengelompokan terhadap data penerimaan mahasiswa baru di Universitas Islam Indragiri. Kluster mahasiswa yang terbentuk adalah tiga kluster, dengan kluster pertama 195 items, kluster kedua 271 items dan kluster ketiga sejumlah 50 items. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi mempromosikan masing-masing program studi yang ada di Universitas Islam Indragiri. Berdasarkan hasil kluster algoritma K-Means dapat dilihat jurusan/program studi yang di minati di masing-masing sekolah [9]. Oleh karena hal tersebut diatas penelitian ini menerapkan metode *data mining* yaitu menggunakan algoritma K-Means *Clustering* untuk mengetahui pola pemilihan program studi pada mahasiswa baru di STMIK Primakara.

### **METODOLOGI PENELITIAN**

Pelaksanaan penelitian penerapan *data mining* untuk mengetahui pola pemilihan program studi di STMIK Primakara menggunakan algoritma *K-Means clustering* ini dimulai pada bulan Januari hingga September 2021. Adapun hasil penelitian berdasarkan tahapan penelitian yang dilakukan dijabarkan sebagai berikut.

#### **1. Identifikasi Masalah**

Dalam tahap ini ditemukan permasalahan yang terjadi adalah banyaknya data yang tersimpan dalam *database* STMIK Primakara belum termanfaatkan dengan baik selain sebagai kebutuhan administrasi. Selain itu pihak perguruan tinggi belum dapat menentukan strategi promosi yang tepat, dimana sasaran/target promosi program studi selama ini menyasar seluruh daerah di Bali dan NTB. Promosi yang dilakukan oleh tim marketing selama ini tidak mengacu pada sebaran peminatan mahasiswa pada setiap program studi di setiap daerahnya. Hal tersebut dianggap kurang efektif dan efisien serta belum memberikan dampak yang signifikan terhadap peningkatan jumlah mahasiswa yang diperoleh.

Oleh sebab itu, dirumuskan tujuan penelitian yakni untuk mengetahui pola pemilihan program studi dan mengetahui pola minat mahasiswa baru terhadap program studi di STMIK Primakara, serta dapat menjadi rujukan pengetahuan agar sosialisasi dan promosi program studi menjadi lebih tepat sasaran.

**2. Studi Literatur**

Studi literatur telah dilakukan diantaranya mengkaji dan menggali pengetahuan mengenai metode *data mining*, teknik *clustering*, algoritma K-Means, serta penelitian terkait yang telah dilakukan melalui buku, jurnal, artikel, *paper*, maupun *internet*.

**3. Pengumpulan Data**

Pada tahap ini dilakukan pengumpulan *dataset* mahasiswa yang didapatkan dari bagian PPTI STMIK Primakara berupa dokumen *excel* data mahasiswa dari tahun 2015 sampai 2020 sejumlah 824 *record*. *Dataset* mahasiswa terdiri dari atribut Nama, Asal Sekolah, Asal Daerah, Tahun Lulus, Jenis Kelamin, dan Program Studi.

**4. Data Preprocessing**

Setelah pengumpulan data, selanjutnya dilakukan tahap *data preprocessing* dimana terdiri dari beberapa tahapan sebagai berikut.

- a) *Data Cleaning*  
 Pada tahap ini dilakukan evaluasi terhadap kualitas data. Data yang tidak memiliki nilai maka data dibuang atau tidak digunakan dalam *dataset*. Hasil dari evaluasi menunjukkan dari total 824 menghasilkan sejumlah 743 data.
- b) *Data Integration*  
 Selanjutnya tahap *data integration* dimana data yang berjenis data nominal diinisialisasikan ke dalam bentuk angka dapat diolah dengan menggunakan algoritma K-Means *clustering*. Namun proses tersebut tidak dilakukan karena atribut data seluruhnya merupakan data kategorikal.
- c) *Data Selection*  
 Pada tahapan ini dilakukan seleksi atribut-atribut data yang dianggap relevan dengan rumusan masalah. Dalam penelitian ini hanya beberapa atribut data saja yang

digunakan yaitu asal daerah, jenis sekolah, dan program studi yang dipilih.

d) *Data Transformation*

Pada tahap ini dilakukan penyesuaian data ke dalam bentuk *.csv* agar data dapat diolah dengan menggunakan *tools data mining* dengan mengimplementasikan algoritma K-Means.

Data yang telah melalui proses *preprocessing* ditunjukkan pada Tabel 1 berikut.

Tabel 1. *Dataset* setelah *Preprocessing*

Kode	Asal Daerah	Jenis Sekolah	Program Studi
1	Semarang	SMA	SI
2	Gianyar	SMK	SIA
3	Gianyar	SMA	SIA
4	Badung	SMK	SIA
5	Denpasar	SMK	SIA
6	Gianyar	SMK	SIA
7	Denpasar	SMA	IF
8	Jembrana	SMA	SI
9	Karangasem	SMK	SI
10	Gianyar	SMK	SI
11	Denpasar	SMK	SI
12	Surakarta	SMK	IF
13	Mataram	SMK	IF
14	Karangasem	SMK	SIA
15	Klungkung	SMA	SI
16	Badung	SMA	SIA
17	Gianyar	SMA	SIA
18	Tabanan	SMA	IF
19	Tabanan	SMA	SI
...			

**5. Penerapan Algoritma K-Means**

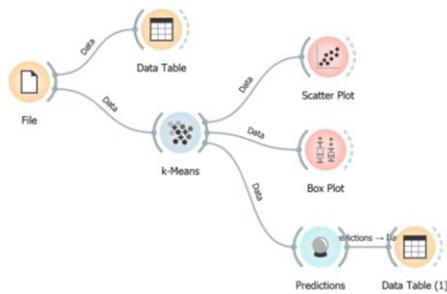
Penerapan algoritma K-Means dilakukan dengan menggunakan *tools data mining* yaitu *Orange Data Mining*.

**6. Data Interpretation/Evaluation**

Pada tahap ini dilakukan penyajian pola atau pengetahuan yang ditemukan pada *dataset* mahasiswa menggunakan algoritma K-Means. Pola ini akan menjadi *output* dari penelitian.

**HASIL DAN PEMBAHASAN**

Penerapan algoritma K-Means dilakukan dengan menggunakan *tools data mining* yaitu *Orange Data Mining*. *Orange data mining* merupakan teknologi pembelajaran mesin yang bersifat *open source*. *Orange* dapat digunakan untuk analisis dan visualisasi data eksploratif. Software ini memiliki fitur khusus untuk kepentingan data mining seperti pemodelan prediktif, sistem rekomendasi dan dapat digunakan untuk penelitian genomik, biomedis, bioinformatika, dan pengajaran [10]. Adapun model *clustering* yang diimplementasikan menggunakan *Orange data mining* yaitu seperti pada Gambar 2 berikut.



Gambar 2. *Modeling K-Means Clustering*

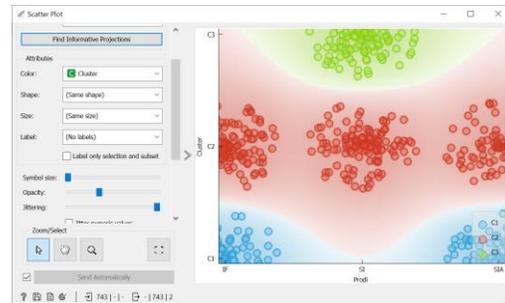
Pada Gambar 2 di atas menunjukkan model penerapan algoritma K-Means *clustering* pada *Orange Data Mining* dengan melibatkan widget *File*, *Data Table*, *k-Means*, *Scatter Plot*, *Box Plot*, dan *Prediction*. Gambar 2 berikut adalah tampilan *dataset* yang telah diinput.

	Daerah	Sekolah	Prodi
1	Ambon	SMK	IF
2	Badung	SMA	IF
7	Badung	SMA	IF
6	Badung	SMA	IF
20	Badung	SMA	IF
21	Badung	SMA	IF
22	Badung	SMA	IF
24	Badung	SMA	IF
26	Badung	SMA	IF
27	Badung	SMA	IF
28	Badung	SMA	IF
30	Badung	SMA	IF
36	Badung	SMA	IF
3	Badung	SMK	IF
4	Badung	SMK	IF
5	Badung	SMK	IF
6	Badung	SMK	IF
9	Badung	SMK	IF
10	Badung	SMK	IF

Gambar 3. *Dataset Table Mahasiswa*

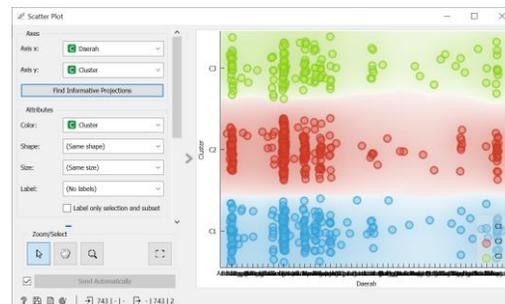
Pada Gambar 3 di atas diketahui bahwa jumlah data yang akan dikelola sebesar 743 data dengan 3 (tiga) *features* atau atribut. Selanjutnya dari model yang telah diterapkan menghasilkan hasil *clustering*. Hasil proses ini

adalah data yang dikelompokkan berdasar kemiripan karakteristik dari setiap data, sehingga dapat ditemukan pola atau informasi yang tersembunyi dari data-data tersebut. *Cluster* yang ditemukan berjumlah 3 (tiga) *cluster* dengan pola seperti yang ditunjukkan pada gambar berikut.



Gambar 4. Hasil *Clustering* hubungan *Cluster* dengan Program Studi

Berdasarkan Gambar 4 di atas menunjukkan bahwa cluster yang terbentuk berjumlah 3 cluster dimana C1 rata-rata berasal dari IF dan SIA, C2 rata-rata berasal dari IF, SI, dan SIA, serta C3 rata-rata berasal dari program studi SI. Selain itu, apabila dilihat dari hubungan antara cluster dengan daerah asal maka pola yang dihasilkan seperti pada Gambar 4 berikut.



Gambar 5. Hasil *Clustering* hubungan *Cluster* dengan Daerah Asal

Dari gambar tersebut dapat diketahui hasil bahwa C1 memiliki 240 data, C2 338 data, dan C3 berjumlah 165 data. Selain di atas, hasil yang diperoleh juga berupa *Box Plot*, seperti yang ditunjukkan pada gambar berikut.



Gambar 6. Hasil Grafik *Box Plot*

Berdasarkan hasil grafik *Box Plot* yang ditunjukkan pada Gambar 5 dapat diketahui bahwa pada C1 diperoleh hasil bahwa mahasiswa yang berada pada *cluster* ini lebih banyak memilih program studi SIA, dimana jenis sekolahnya adalah SMA dengan 30.83% berasal dari Denpasar, 15% mahasiswa berasal dari Badung, 10.42% dari Gianyar, 7.50% dari Tabanan, 6.25% berasal dari Jemberana. Sedangkan pada C2 diperoleh hasil bahwa mahasiswa yang berada pada *cluster* ini lebih banyak memilih program studi IF dimana jenis sekolahnya adalah SMK dengan 45.27% berasal dari Denpasar, 15.98% berasal dari Badung, 13.31% dari Gianyar, 6.51% berasal dari Karangasem, dan 5.33% berasal dari Tabanan. Serta pada C3 diperoleh hasil bahwa mahasiswa yang berada pada *cluster* ini seluruhnya memilih program studi SI, dimana jenis sekolahnya adalah SMA dengan 30.91% berasal dari Denpasar, 15.76% berasal dari

Badung, 8.70% berasal dari Gianyar, serta masing-masing 6.06% berasal dari Jemberana dan Tabanan.

## 7. Data Interpretation/Evaluation

Pada tahap ini dilakukan penyajian pola atau pengetahuan yang ditemukan pada *dataset* mahasiswa menggunakan algoritma K-Means. Pola ini akan menjadi *output* dari penelitian. *Output* tersebut akan menjawab masalah yang telah dirumuskan di awal karena pada proses ini dilakukan formulasi keputusan atau aksi dari hasil yang didapat. Berikut adalah pola yang dihasilkan dalam setiap *cluster*.

### a. Cluster 1 (C1)

*Cluster 1* terdiri dari mahasiswa yang lebih banyak memilih program studi SIA berasal dari sekolah SMA di daerah Denpasar, Badung, Gianyar, Tabanan, dan Jemberana.

### b. Cluster 2 (C2)

*Cluster 2* terdiri dari mahasiswa yang lebih banyak memilih program studi IF berasal dari sekolah SMK di daerah Denpasar, Badung, Gianyar, Karangasem, dan Tabanan.

### c. Cluster 3 (C3)

*Cluster 3* terdiri dari mahasiswa yang lebih banyak memilih program studi SI berasal dari sekolah SMA di daerah Denpasar, Badung, Gianyar, Tabanan, dan Jemberana.

Pola *clustering* yang dihasilkan pada penelitian ini selanjutnya diberikan kepada pihak pemangku kepentingan di STMIK Primakara untuk dapat dijadikan acuan/pedoman dalam menentukan kebijakan dalam upaya mendukung strategi promosi yang tepat dalam mendapatkan calon mahasiswa baru.

## SIMPULAN

Setelah dilakukan pengelompokan data mahasiswa berdasarkan program studi, asal daerah, dan jenis sekolah menggunakan K-Means *clustering* terbentuk 3 (tiga) *cluster*. *Cluster 1* lebih banyak memilih program studi SIA berasal dari sekolah SMA di daerah

Denpasar, Badung, Gianyar, Tabanan, dan Jembrana. *Cluster 2* yang lebih banyak memilih program studi IF berasal dari sekolah SMK di daerah Denpasar, Badung, Gianyar, Karangasem, dan Tabanan. *Cluster 3* terdiri dari mahasiswa yang lebih banyak memilih program studi SI berasal dari sekolah SMA di daerah Denpasar, Badung, Gianyar, Tabanan, dan Jembrana. Hasil *clustering* ini selanjutnya dapat digunakan untuk menentukan strategi promosi pada persebaran daerah berdasarkan jenis sekolah dan program studi yang diminati.

#### DAFTAR PUSTAKA

- [1] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2017.
- [2] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [3] S. Rony, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru ( Studi Kasus : Politeknik Lp3i Jakarta )," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.
- [4] R. Primartha, *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika, 2018.
- [5] B. Poerwanto and R. Y. Fa'rifah, "Analisis Cluster K-Means dalam Pengelompokan Kemampuan Mahasiswa," *J. Sci. Pini*, vol. 2, no. 2, pp. 92–96, 2016.
- [6] K. Sya'iyah, H. Yuliansyah, and I. Arfiani, "Clustering student data based on K-means algorithms," *Int. J. Sci. Technol. Res.*, vol. 8, no. 8, pp. 1014–1018, 2019.
- [7] O. Tinuke Omolewa, A. Taye Oladele, A. Adekanmi Adeyinka, and O. Roseline Oluwaseun, "Prediction of Student's Academic Performance using k-Means Clustering and Multiple Linear Regressions," *J. Eng. Appl. Sci.*, vol. 14, no. 22, pp. 8254–8260, 2019, doi: 10.36478/jeasci.2019.8254.8260.
- [8] E. T. Munawaroh, R. I. Handayani, and E. Widanengsih, "The Implementation of K-Means Algorithm for Cluster Majoring to New Students in SMKN 2 of South Tangerang," *Sinkron*, vol. 4, no. 1, p. 164, 2019, doi: 10.33395/sinkron.v4i1.10133.
- [9] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [10] R. A. raffaidy Wiguna and A. I. Rifai, "Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021, doi: 10.33557/journalisi.v3i1.78.