

## DATA MINING STRATEGI PROMOSI PADA UNIVERSITAS YAPIS PAPUA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Mursalim Tonggiroh<sup>1</sup>, Muhammad Taher Jufri<sup>2</sup>.

Fakultas Teknik dan Sistem Informasi

Universitas Yapis Papua

email: <sup>1</sup>mursalim.t@gmail.com, <sup>2</sup>jufri.conoras@gmail.com

Jl. Sam Ratulangi No. 11, Jayapura. Papua. 99115, Indonesia

### Abstrak

Universitas Yapis Papua setiap tahun melakukan penerimaan mahasiswa baru, hal ini akan berdampak terhadap pertumbuhan jumlah data mahasiswa yang sangat pesat. Dengan jumlah data mahasiswa yang besar terdapat informasi yang tersembunyi yang dapat diketahui dengan cara melakukan pengolahan data tersebut sehingga berguna bagi pihak Universitas. Penelitian ini akan dilakukan dengan menggunakan algoritma K-means Clustering. Dengan menggunakan algoritma ini, berbagai data mahasiswa yang telah dikumpulkan akan dikelompokkan ke dalam beberapa cluster. Atribut yang akan digunakan dalam penentuan pengelompokan wilayah untuk melakukan promosi UNYAP adalah NPM, program studi, asal distrik dan status mahasiswa. Hasil dari penelitian ini memberikan strategi promosi penerimaan mahasiswa baru berdasarkan dari analisa data mahasiswa pada tahun 2015, 2016, 2017. Serta menghasilkan sebuah aplikasi data mining yang membantu pihak Universitas untuk menganalisa data tersebut.

Kata kunci: Data Mining, *K-means Clustering*, Strategi Promosi, Universitas Yapis Papua

## 1 PENDAHULUAN

### 1.1 Latar Belakang

Institusi pendidikan seperti Perguruan Tinggi atau Universitas akan mengalami peningkatan jumlah data setiap tahun seperti data pendaftaran calon mahasiswa baru, data mahasiswa, data nilai, dan data alumni. Data akan terus menumpuk dan menjadi kuburan data serta tidak memiliki informasi yang dapat dihasilkan dari data tersebut. Oleh sebab itu diperlukan suatu teknik atau perangkat untuk membantu dalam mentransformasikan data dengan jumlah yang besar tersebut menjadi informasi berguna yaitu dengan penerapan dengan data mining untuk strategi promosi penerimaan mahasiswa baru yang dapat menjadi dasar atau pedoman dalam mendukung pengambilan sebuah keputusan, terutama keputusan tentang strategi promosi dalam penerimaan mahasiswa baru.

Universitas Yapis Papua (UNYAP) merupakan salah satu perguruan tinggi swasta yang di kota Jayapura. UNYAP setiap tahun melakukan penerimaan mahasiswa baru, hal ini akan berdampak terhadap pertumbuhan jumlah data mahasiswa yang sangat pesat. Dengan jumlah data mahasiswa yang besar terdapat informasi yang tersembunyi yang dapat diketahui dengan cara melakukan pengolahan data tersebut sehingga berguna bagi pihak Universitas. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge discovery*), misalnya informasi mengenai pengelompokan data mahasiswa berpotensi berdasarkan asal distrik mahasiswa. Pengetahuan baru tersebut dapat membantu pihak Universitas dalam menentukan strategi promosi terhadap calon mahasiswa baru dengan tepat sasaran.

Selama ini UNİYAP melakukan strategi promosi seperti sosialisasi ke SMA/SMK yang ada di wilayah Jayapura ataupun promosi melalui media cetak dan media massa. Wilayah promosi yang dilakukan dari tahun ke tahun meningkat. Sedangkan selama 3 tahun terakhir jumlah mahasiswa baru mengalami kenaikan dan penurunan seperti pada tahun 2015 pendaftar berjumlah 887 orang. Tahun 2016 berjumlah 1014 orang, sedangkan tahun 2017 berjumlah 941 orang. Berdasarkan data yang telah diperoleh dari tahun 2015 ke 2016 mahasiswa baru mengalami peningkatan berjumlah 127 orang. Berbeda dari tahun sebelumnya pada tahun 2017 mengalami penurunan mahasiswa baru yaitu berjumlah 73 orang. Salah satu penyebab terjadinya penurunan jumlah mahasiswa pada tahun 2017 adalah kurang dilakukan pengolahan data mahasiswa baru secara tepat berdasarkan data historis oleh pihak Biro Informasi, Penerimaan dan Pengembangan Institusi (BIP2I) UNİYAP. Hal tersebut dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Selain itu menentukan wilayah promosi yang tepat sasaran, dampaknya adalah dapat meminimalisir biaya promosi mahasiswa baru yang akan dilakukan. Oleh karena itu, pengolahan data mahasiswa baru seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya. Penelitian ini akan dilakukan dengan menggunakan algoritma *K-means Clustering*. Dengan menggunakan algoritma ini, berbagai data mahasiswa yang telah dikumpulkan akan dikelompokkan ke dalam beberapa *cluster*. Atribut yang akan digunakan dalam penentuan pengelompokan wilayah untuk melakukan promosi UNİYAP adalah NPM, program studi, asal distrik dan status mahasiswa. Hasil dari pengelompokan data tersebut bertujuan untuk membantu pihak Universitas dalam melakukan promosi dengan strategi yang tepat untuk meningkatkan jumlah mahasiswa baru tiap tahunnya. Maka ada rumusan permasalahan yang dapat yaitu bagaimana menerapkan data mining dengan menggunakan algoritma *K-means Clustering* dalam menentukan strategi promosi penerimaan mahasiswa baru yang tepat sasaran.

Penelitian ini bertujuan untuk mengimplementasikan konsep *data mining* dengan menggunakan Algoritma K-means dalam mengetahui persebaran wilayah distrik berdasarkan data mahasiswa tahun 2015, 2016, dan 2017 berdasarkan potensi status mahasiswa yang aktif dan menentukan strategi promosi yang tepat untuk setiap *cluster* yang terbentuk.

## 1.2 Penelitian Terdahulu

Penelitian Rima Dias Ramadhani melakukan *clustering* data mahasiswa baru dengan algoritma *k-means*. Atribut yang digunakan adalah kota asal, program studi dan IPK mahasiswa yang dibagi menjadi 3 *cluster*. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan *cluster* yang terbentuk oleh pihak akademisi UDINUS (Ramadhani, 2014). Johan Oscar Ong melakukan penelitian untuk menentukan strategi marketing President University. Dalam penelitian ini, melakukan perhitungan *dataset* mahasiswa yang telah lulus dari President University menggunakan algoritma *k-means clustering*. Pengelompokan data mahasiswa ke dalam beberapa kelompok berdasarkan karakteristik dari *dataset* untuk menemukan informasi yang tersembunyi dari *dataset* mahasiswa yang telah lulus dari President University. Data atribut yang digunakan dalam penelitian ini adalah asal kota, program studi dan IPK. Tujuan dari penelitian ini adalah untuk membantu departemen pemasaran President University dalam memprediksi strategi promosi yang dilakukan di kota-kota di Indonesia. Informasi yang diperoleh dalam penelitian ini dapat digunakan sebagai referensi dalam menentukan strategi yang tepat untuk tim pemasaran dalam kegiatan promosi di kota-kota di Indonesia sehingga promosi akan lebih efektif dan efisien (Ong, 2013).

Penelitian berjudul Penerapan Data Mining dengan Algoritma Fp-Growth untuk mendukung Strategi Promosi Pendidikan. Penelitian dilakukan dengan mengamati beberapa variabel penelitian yang sering dipertimbangkan oleh perguruan tinggi khususnya bagian marketing dalam menentukan sasaran promosinya yaitu pendidikan terakhir, alamat rumah, jurusan, pilihan program studi. Hasil penelitian ini adalah berupa suatu perangkat lunak dengan mengimplementasikan algoritma FP-Growth yang menggunakan konsep pembangunan FP-Tree dalam mencari *frequent itemset* (Ikhwan dkk, 2015).

## 2 METODOLOGI PENELITIAN

### 2.1 Metodologi CRISP-DM

Analisis data yang dilakukan sesuai dengan tahapan metodologi CRISP-DM yang telah dicantumkan pada landasan teori. Berikut tahapan yang akan dilakukan dalam menganalisis data yaitu:

1. Pemahaman Bisnis (*Business Understanding*), merupakan tahap pertama dalam proses CRISP-DM yang juga dapat disebut tahap pemahaman bisnis (penelitian).
  - a. Menentukan Tujuan Bisnis, tujuan bisnis berdasarkan pengolahan data mahasiswa tahun 2015, 2016 dan 2017 antara lain untuk meningkatkan mahasiswa pada tahun berikutnya. Pengolahan data dilakukan karena telah terjadi penurunan sejumlah 73 mahasiswa pada tahun 2017. Keputusan untuk melakukan wilayah promosi yang tepat sasaran.
  - b. Menentukan strategi awal *data mining*, strategi awal dalam menerapkan tujuan dilakukannya data mining adalah melakukan permintaan data terlebih dahulu ke bagian Pusat Data (P-Data) UNIYAP.
2. Pemahaman Data (*Data Understanding*), *dataset* mahasiswa baru yang didapatkan dari P-Data UNIYAP berupa dokumen excel sejumlah 2.995 data mahasiswa.
  - a. Pengumpulan data awal, adapun sumber utama yang digunakan dalam penelitian ini adalah *dataset* mahasiswa baru UNIYAP tahun 2015, 2016 dan 2017.
  - b. Mendeskripsikan data, *dataset* mahasiswa terdiri dari atribut NPM, nama, program studi, alamat, asal sekolah, jenis kelamin, tempat tanggal lahir, tahun masuk, tahun masuk SMA, nomor telepon. Jumlah data yang ada pada atribut berjumlah 70.608 *record*.
  - c. Pemilihan Atribut, atribut yang digunakan adalah NPM, program studi mahasiswa, asal distrik mahasiswa, dan status mahasiswa.
3. Persiapan Data (*Data Preparation*), pada bagian ini data akan dipersiapkan agar lebih mudah dalam proses *mining*. Proses *preparation* akan mencakup 3 hal mendasar yaitu:
  - a. Seleksi data, seleksi data merupakan suatu pemilihan himpunan data yang nantinya digunakan dalam penulisan yaitu data NPM, Program Studi, alamat, dan status mahasiswa.
  - b. *Cleaning*, pada proses ini data akan dibersihkan meliputi kelengkapan data, penghapusan data *redundant* (data yang sama dalam *database*) dan penghilangan *noise*.
  - c. *Transformation*, data yang berjenis nominal seperti program studi, asal distrik, dan status mahasiswa harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk angka.

Untuk melakukan inialisasi dapat dilakukan dengan:

1. Inialisasi asal distrik yang dilakukan dengan membagi nama distrik yang ada di Kota Jayapura. Dapat dilihat pada tabel 1 berikut.

Tabel 1. Inialisasi Asal Distrik

Asal Distrik	Frekuensi	Inisial
Jayapura Utara	1.249	1
Jayapura Selatan	651	2
Abepura	221	3
Heram	70	4
Lain-lain	61	5
Muara Tami	7	6

2. Inialisasi Program Studi dilakukan berdasarkan frekuensi atau jumlah mahasiswa yang memilih prodi tersebut.

Tabel 2. Inialisasi Program Studi

Program Studi	Frekuensi	Inisial
Manajemen	531	1
Akuntansi	439	2
Ilmu Hukum	402	3
Teknik Sipil	331	4

Program Studi	Frekuensi	Inisial
Sistem Informasi	264	5
Ilmu Pemerintahan	138	6
Administrasi Negara	83	7
Pendidikan Agama Islam	40	8
Perikanan	31	9

3. Inisialisasi status mahasiswa dibagi menjadi dua jenis yaitu aktif dan non aktif.

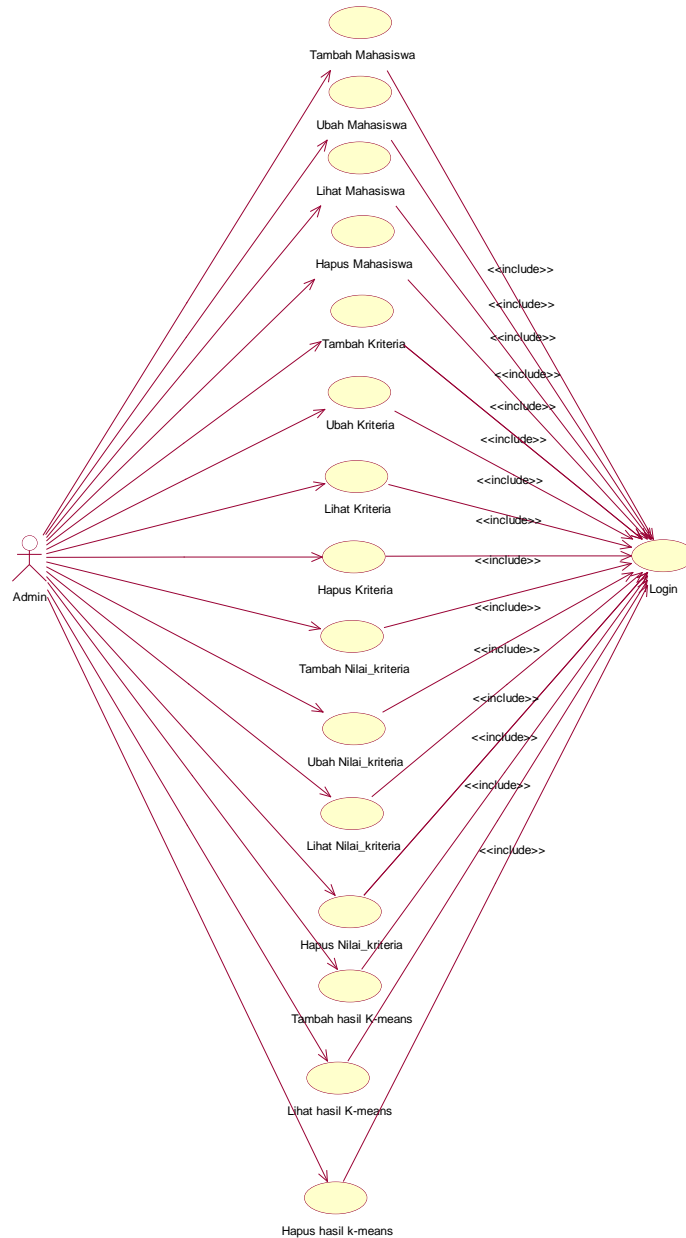
Tabel 3. Inisialisasi Status Mahasiswa

Status Mahasiswa	Frekuensi	Inisial
Aktif	1.893	0
Non Aktif	366	1

4. Pemodelan (*Modelling*), teknik pemodelan yang digunakan adalah *clustering* menggunakan algoritma *K-Means*.
5. Evaluasi (*evaluation*), tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan data mining yang telah ditentukan pada tahap *business understanding*. Pada tahap ini juga harus dipastikan bahwa semua tahapan/faktor penting yang telah dilakukan dalam pengolahan data tidak ada yang terlewatkan.
6. Penyebaran (*deployment*), proses data mining yang dihasilkan disampaikan kepada pihak BIP2I sebagai rekomendasi untuk menentukan strategi promosi yang digunakan pada penerimaan mahasiswa baru.

## 2.2 Perancangan Sistem

Perancangan sistem menggunakan *Unified Modeling Language* (UML). Salah satu diagram yang digunakan yaitu *Use Case Diagram*. Terdapat satu aktor yaitu admin yang menggunakan sistem.



Gambar 1: Use Case Diagram sistem

Gambar 2 menunjukkan *use case diagram* sistem. Aktor admin harus *login* terlebih dahulu. Setelah *login* admin dapat menambahkan mahasiswa, melihat mahasiswa, mengubah mahasiswa, menghapus mahasiswa, menambahkan kriteria, melihat kriteria, mengubah kriteria menghapus kriteria, menambahkan nilai kriteria, melihat nilai kriteria, mengubah nilai kriteria, menghapus nilai kriteria.

### 3 HASIL DAN PEMBAHASAN

#### 3.1 Implementasi Algoritma K-Means Clustering

Berikut penerapan algoritma *K-means* pada data mahasiswa yang telah dilakukan inisialisasi guna untuk *clustering* strategi promosi yang tepat sasaran.

Tabel 4. Contoh *Dataset* Mahasiswa yang telah dilakukan Inisialisasi

NPM	Program Studi	Asal Distrik	Status Mahasiswa
15121038	2	2	0
15111092	1	1	1
15121060	2	3	0

NPM	Program Studi	Asal Distrik	Status Mahasiswa
16311073	3	1	1
17621075	4	2	0
15611021	5	2	0
15221027	6	5	1
15211022	7	2	0
15411002	9	5	0
16411003	9	3	1

Tabel 4 menunjukkan contoh dataset mahasiswa sebanyak 10 mahasiswa dari jumlah data yang dikumpulkan yaitu 2.259 mahasiswa.

1. Penetapan jumlah *cluster* (K) pada penelitian ini yaitu berjumlah 3 *cluster*. Berikut tabel titik pusat awal tiap *cluster*.

Tabel 5. Titik Pusat Awal *Cluster*

Titik Pusat Awal <i>Cluster</i>	NPM	Program Studi	Asal Distrik	Status Mahasiswa
<i>Cluster 1</i>	15121038	2	2	0
<i>Cluster 2</i>	16311073	3	1	1
<i>Cluster 3</i>	17621075	4	2	0

2. Selanjutnya menghitung jarak terpendek dengan menggunakan rumus *Euclidean*.  
 Jarak data dengan *cluster 1* yaitu:  $D_{x_{15121038}, C1} = \sqrt{(2-2)^2 + (2-2)^2 + (0-0)^2} = 0,000$   
 Jarak data dengan *cluster 2* yaitu:  $D_{x_{15121038}, C2} = \sqrt{(2-3)^2 + (2-1)^2 + (0-1)^2} = 1,732$   
 Jarak data dengan *cluster 3* yaitu:  $D_{x_{15121038}, C3} = \sqrt{(2-4)^2 + (2-2)^2 + (0-0)^2} = 2,000$
3. Setelah perhitungan jarak terpendek selesai, langkah selanjutnya kelompokkan data berdasarkan jarak terpendek. Contoh: mahasiswa dengan NPM 15511001 memiliki jarak terpendek pada *cluster 2* maka jarak terdekat adalah 2. Hasil perhitungan iterasi-1 dapat dilihat pada tabel 6 berikut:

Tabel 6. Hasil Perhitungan Jarak Iterasi-1

NPM	Jarak Ke			Jarak Terdekat
	C1	C2	C3	
15121038	0,00	1,73	2,00	C1
15111092	1,73	2,00	3,32	C1
15121060	1,00	2,45	2,24	C1
16311073	1,73	0,00	1,73	C2
17621075	2,00	1,73	0,00	C3
15611021	3,00	2,449	1,00	C3
15221027	5,09	5,00	3,74	C3
15221022	5,00	4,24	3,00	C3
15411002	7,61	7,28	5,83	C1
16411003	7,14	6,33	5,20	C3

4. Tentukan pusat *centroid* baru dengan cara menghitung rata-rata dari keanggotaan mahasiswa
5. Setelah mendapatkan pembaruan *centroid*, langkah selanjutnya ulangi iterasi dengan pusat *centroid* baru dengan cara perhitungan yang sama dengan iterasi 1 seperti langkah 2 sampai 5 hingga tidak ada data yang berpindah.
6. Telah didapat iterasi-4 menunjukkan bahwa sudah tidak adanya perpindahan data dengan iterasi sebelumnya dengan demikian iterasi tersebut dapat dinyatakan berhenti karena telah memenuhi kriteria Algoritma *K-means*.
7. Selanjutnya melakukan analisa hasil *cluster* yang telah dilakukan. Analisa hasil *cluster* dapat dilihat pada tabel 7, tabel 8 dan tabel 9.

Tabel 7. Hasil Analisis *Cluster 1*

<b>Cluster 1 terdiri dari 970 mahasiswa yang berasal dari:</b>	
Program Studi: a. Manajemen =531 b. Akuntansi = 439	Asal Distrik: a. Jayapura Utara = 527 b. Jayapura Selatan = 315 c. Abepura = 91 d. Heram = 21

e. Lain-lain = 14 f. Muara tami = 2
Status Mahasiswa: Aktif = 840 Non = 130 Presentasi status mahasiswa aktif: $840/2259 \times 100\% = 37,18\%$

Tabel 8. Hasil Analisis *Cluster 2*

<b>Cluster 2 terdiri dari 660 mahasiswa yang berasal dari:</b>	
Program Studi: a. Ilmu Hukum = 402 b. Akuntansi = 439	Asal Distrik: a. Jayapura Utara = 402 b. Jayapura Selatan = 173 c. Abepura = 55 d. Heram = 23 e. Lain-lain = 7
Status Mahasiswa: Aktif = 551 Non = 109 Presentasi status mahasiswa aktif: $551/2259 \times 100\% = 24,39\%$	

Tabel 9. Hasil Analisis *Cluster 3*

<b>Cluster 3 terdiri dari 629 mahasiswa yang berasal dari:</b>	
Program Studi: a. Teknik Sipil = 335 b. Sistem Informasi = 2 c. Ilmu pemerintahan = 138 d. Administrasi Negara = 83 e. Pendidikan Agama Islam = 40 f. Perikanan = 31	Asal Distrik: a. Jayapura Utara = 320 b. Jayapura Selatan = 163 c. Abepura = 75 d. Heram = 26 e. Lain-lain = 40 f. Muara Tami = 5
Status Mahasiswa: Aktif = 502 Non = 127 Presentasi status mahasiswa aktif: $502/2259 \times 100\% = 22,22\%$	

### 3.2 Strategi Promosi

Dari data hasil *clustering* yang telah dilakukan di atas, maka dapat ditentukan beberapa strategi promosi yang dapat dilakukan oleh pihak BIP2I.

1. Promosi dengan mengirim tim BIP2I yang sesuai dengan program studi yang paling banyak diminati.
2. Promosi pada persebaran wilayah distrik berdasarkan persentase status mahasiswa aktif.

Selanjutnya dilakukan analisis strategi promosi berdasarkan *promotion mix* pada masing-masing *cluster* yang terbentuk. Diharapkan dengan strategi yang diberikan pada masing-masing *cluster* dapat sesuai dengan tujuan bisnis yang telah dijelaskan.

Tabel 10. Strategi Promosi berdasarkan *Promotion Mix*

No	Strategi Promosi	Cluster 1	Cluster 2	Cluster 3
1	Periklanan	√	√	√
2	Penjualan Personal	√	√	√
3	Promosi Penjualan		√	√
4	Hubungan Masyarakat		√	√

### 3.3 Tampilan Sistem Analisis K-Means

Halaman ini digunakan untuk menghitung proses analisis *K-means Clustering*. Halaman ini terdapat *form* untuk memasukkan jumlah *cluster* dan maksimum iterasi.

NPM	Kriteria		
	Predi	Asal Distrik	Status_Mhs
14020020	1	2	1

Jumlah Cluster:

Maksimum Iterasi:

Activate Windows  
Go to Settings to activate Windows

Gambar 2: Tampilan halaman analisis K-Means

Pengujian yang dilakukan pada sistem seperti pengujian *interface* sistem, pengujian fungsi dasar sistem dan pengujian *form handle* sistem. Hasil yang didapatkan pada pengujian tersebut yaitu semua pengujian yang dilakukan dinyatakan sukses

#### 4 KESIMPULAN

Berdasarkan hasil dan pembahasan maka dapat disimpulkan sebagai berikut:

1. Setelah dilakukan pengelompokan data mahasiswa melalui persebaran distrik di Kota Jayapura berdasarkan presentasi status mahasiswa aktif menggunakan *K-means Clustering* terbentuk tiga *cluster* yaitu, *cluster* satu dengan jumlah mahasiswa 1.006 mahasiswa dengan persentase mahasiswa aktif 37,5%, *cluster* dua dengan jumlah mahasiswa 650 mahasiswa dengan persentase mahasiswa aktif 23,32%, *cluster* tiga dengan jumlah mahasiswa 655 mahasiswa dengan persentase mahasiswa aktif 22,76%.
2. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah distrik di Kota Jayapura berdasarkan *cluster* yang terbentuk adalah dengan mengirim tim BIP2I yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan persentase status mahasiswa aktif dengan melakukan penyesuaian menggunakan *promotion mix* dan dengan melihat persentase status mahasiswa aktif pada setiap *cluster*.

#### Referensi

- Ikhwan, A., Nofriansyah, D., dan Sriani., (2015), 'Penerapan Data Mining dengan Algoritma FP-Growth Untuk Mendukung Strategi Promosi Pendidikan (Studi kasus Kampus STMIK Triguna Darma'. *Jurnal Ilmiah SAINTIKOM*, vol. 14, no. 3, hh 212-214.
- Ong, J. O., (2013), 'Implementasi Algoritma K-means Clustering Untuk Menentukan Strategi Marketing President University'. *Jurnal Ilmiah Teknik Industri*, vol. 02, no. 01, hh 10-11.
- Ramadhani, R. D., (2014), 'Data Mining Menggunakan Algoritma K-means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro'. *Jurnal Sistem Informasi*, hh. 2-3.