KAJIAN *DATA MINING* MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM STRATEGI PROMOSI

(Studi Kasus: UNISA Kuningan)

TESIS

Disusun sebagai salah satu syarat untuk memperoleh gelar Magister Komputer
dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh:

NENG SRI LATHIFAH ZULFA

NPM : 2018210052

PROGRAM STUDI PASCASARJANA MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG

2020

KAJIAN *DATA MINING* MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM STRATEGI PROMOSI

(Studi Kasus: UNISA Kuningan)

Oleh:

NENG SRI LATHIFAH ZULFA

NPM : 2018210052

Bandung, 28 Maret 2020
Menyetujui,

Dr. Eng. H. Ana Hadiana
Pembimbing

PROGRAM STUDI PASCASARJANA MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG

2020

KATA PENGANTAR

Alhamdulillah, segala puji syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan ridho-NYA, sehingga tesis dengan judul "Kajian Data Mining Menggunakan Algoritma *K-Means* dan *K-Medoids* Dalam Strategi Promosi" ini dapat diselesaikan. Penulis lebih khusus mengucapkan terimakasih kepada Bapak Dr. Eng. H. Ana Hadiana atas bimbingan, arahan dan waktu yang telah diluangkan kepada penulis untuk berdiskusi selama menjadi dosen pembimbing sehingga akhirnya tesis ini dapat selesai.

Penulis menyadari bahwa tesis ini dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis juga ingin menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya, kepada :

1. Nurul Iman Hima Amrullah, S.Ag., M.Si. selaku Rektor Universitas Islam Al-Ihya (UNISA) Kuningan yang telah memberikan kesempatan kepada penulis untuk melaksanakan penelitian di Universitas Islam Al-Ihya (UNISA) Kuningan.
2. Seluruh Civitas Akademika Universitas Islam Al-Ihya (UNISA) Kuningan yang telah bekerja sama dalam proses pengumpulan data dan wawancara yang tidak dapat penulis sebutkan satu persatu.
3. Seluruh Civitas Akademika STMIK LIKMI yang telah banyak membantu kelancaran aktifitas perkuliahan.
4. Kedua Orang Tua penulis yang selalu mendo'akan, memperjuangkan, mendidik serta memberikan dukungan baik moril maupun materil.
5. Suamiku yang telah mendo'akan, membantu dan memberikan dukungan.
6. Rekan rekan seperjuangan program Pascasarjana (S2) angkatan 2018 yang tidak dapat penulis sebutkan satu persatu.
7. Kepada semua pihak yang telah membantu yang tidak dapat penulis sebutkan satu persatu.

8. Dengan keterbatasan pengalaman, ilmu maupun pustaka yang ditinjau, penulis menyadari bahwa tesis ini masih banyak kekurangan dan pengembangan lanjut agar benar benar bermanfaat. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran agar tesis ini lebih sempurna serta sebagai masukan bagi penulis untuk penelitian dan penulisan karya ilmiah di masa yang akan datang.

Bandung, Maret 2020

Penulis

ABSTRAK

KAJIAN *DATA MINING* MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM STRATEGI PROMOSI (Studi Kasus: UNISA Kuningan)

Oleh :

Nama : Neng Sri Lathifah Zulfa
NPM : 2018210052

Proses penerimaan calon mahasiswa baru di UNISA Kuningan dilakukan setiap tahun sehingga menghasilkan data yang banyak berupa profil calon mahasiswa. Kegiatan tersebut menimbulkan penumpukan data sehingga menjadi kesulitan dalam melakukan identifikasi terhadap calon mahasiswa. Penelitian ini bertujuan untuk memberikan rekomendasi strategi promosi untuk UNISA Kuningan karena pada tahun 2016 dan 2019 jumlah mahasiswa baru mengalami penurunan. Hal itu bisa dikarenakan strategi promosi yang belum tepat. Selain itu belum adanya penelitian di UNISA Kuningan untuk mendapatkan informasi dalam melakukan strategi promosi dengan cara mengelompokkan data penerimaan mahasiswa baru. Sehingga pada penelitian ini akan memanfaatkan proses *data mining* dengan teknik *clustering* untuk mendapatkan rekomendasi strategi promosi yang sesuai.

Peneltian ini menggunakan data mahasiswa baru tahun 2014 sampai dengan 2019 yang berjumlah 2.047 *record* dengan metodologi *Knowledge Discovery In Databases* (KDD) sebagai tahapan pemecahan masalah untuk *data mining*. Penelitian ini akan membandingkan algoritma *K-Means* dan *K-Medoids* untuk menghasilkan profil yang memiliki kemiripan atribut yang sama. Validasi Dunn I*ndex* digunakan untuk menentukan K terbaik diantara kedua algoritma tersebut. Implementasi proses menggunakan *Rapid Miner 9.6*.

Berdasarkan hasil implementasi diperoleh *K-Means* sebagai algoritma terbaik dengan nilai validasi dunn index 0,999 yang menghasilkan lima *cluster* sebagai rekomendasi strategi promosi berdasarkan promotion mix.

Kata Kunci : Data Mining, Algoritma, K-Means, Algortma K-Medoids, Clustering, Strategi Promosi, Dunn Index, Rapidminer

*A*BSTRACT

DATA MINING STUDY USING *K-MEANS* AND *K-MEDOIDS* ALGORITHM IN PROMOTIONAL STRATEGY (Case Study: Kuningan UNISA)

By:

Nama : Neng Sri Lathifah Zulfa
NPM : 2018210052

The admission process for new students at UNISA Kuningan is carried out every year so as to produce a lot of data in the form of profiles of prospective students. These activities cause the accumulation of data so that it becomes difficult in identifying prospective students. This study aims to provide a promotional strategy recommendation for UNISA Kuningan because in 2016 and 2019 the number of new students has decreased. That could be due to the inappropriate promotion strategy. In addition, there is no research at UNISA Kuningan to obtain information in conducting promotional strategies by grouping new student admission data. So that in this study will utilize the data mining process with clustering techniques to obtain appropriate promotional strategy recommendations.

This research uses new student data from 2014 to 2019, amounting to 2,047 records with the Knowledge Discovery In Databases (KDD) methodology as a problem solving step for data mining. This study will compare the K-Means and K-Medoids algorithms to produce profiles that have the same attribute. Dunn Index validation is used to determine the best K between the two algorithms. Implementation of the process using Rapid Miner 9.6.

Based on the implementation results obtained by K-Means as the best algorithm with a validation value of 0.999 dunn index which produces five clusters as a recommendation for a promotional strategy based on the promotion mix.

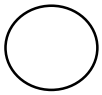
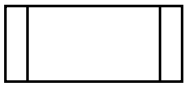
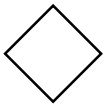
Keywords: Data Mining, Algorithms, K-Means, Algortma K-Medoids, Clustering, Promotion Strategy, Dunn Index, Rapidminer

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SIMBOL

SIMBOL	KETERANGAN
	Terminal (Mulai, Selesai)
	*Input / Output*
	Proses
	*Decision* (Ya, Tidak)
	Alur Proses
	Dokumen

BAB I

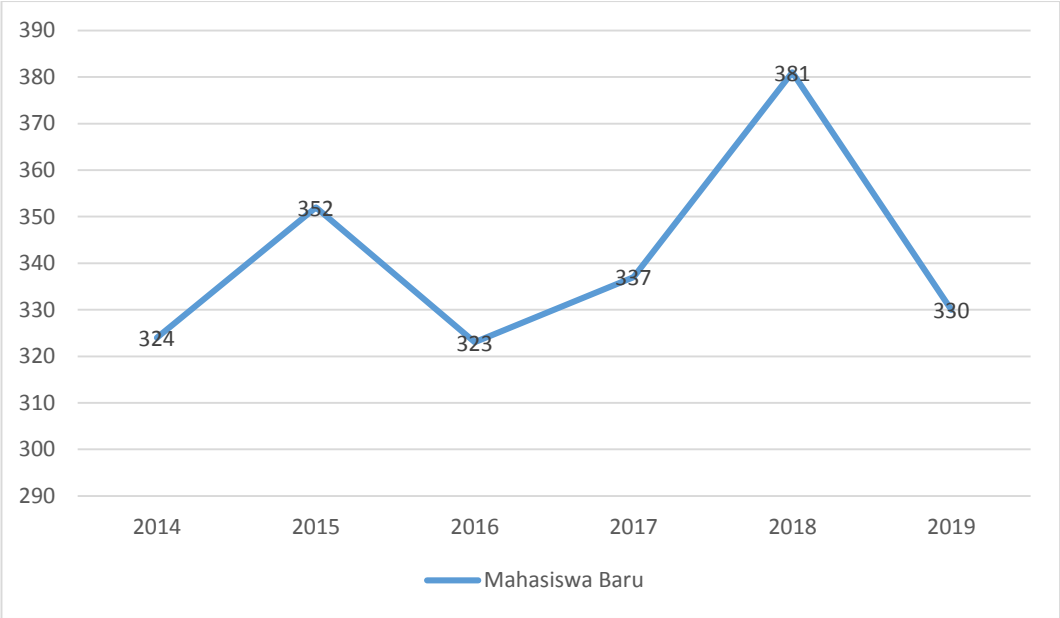
PENDAHULUAN

1.1 Latar Belakang

Kemajuan dan perkembangan teknologi informasi yang pesat, dapat memudahkan dalam proses pencarian informasi maupun dalam menghasilkan informasi seperti pada bidang ekonomi, industri, teknologi, termasuk bidang pendidikan di perguruan tinggi. Penerapan teknologi pada perguruan tinggi dapat diperoleh dari data historis, data akan terus bertambah sehingga terjadi penumpukan data dan dapat memperlambat pencarian informasi terhadap data tersebut. Pengolahan terhadap data yang terus menumpuk dapat di lakukan untuk mengetahui informasi yang tersembunyi.

Universitas Islam Al-Ihya (UNISA) Kuningan merupakan perguruan tinggi swasta di Kuningan Jawa Barat yang harus berusaha dalam meningkatkan jumlah penerimaan mahasiswa baru pada setiap tahun ajaran baru, karena jumlah mahasiswa dapat mempengaruhi kegiatan pembelajaran, operasional dan pengembangan bagi universitas. Calon mahasiswa dan mahasiswa pada umumnya sangat berpengaruh bagi universitas swasta karena uang pendidikan yang diambil dari calon mahasiswa dan mahasiswa digunakan untuk pembiayaan operasional kegiatan belajar mengajar.

Setiap tahun UNISA KUNINGAN melakukan penerimaan mahasiswa baru. Jumlah mahasiswa baru di UNISA Kuningan dari tahun 2014 sampai tahun 2019 seperti pada Gambar 1.1.



Gambar 1.1
Grafik Mahasiswa Baru Tahun 2014-2019 (AIS UNISA, 2019)

Pada Gambar 1.1 terlihat bahwa jumlah mahasiswa mengalami penurunan pada tahun 2016 dan 2019. Selain itu, berdasarkan wawancara dengan bagian Pusat Informasi (PUSINFO) UNISA Kuningan ada beberapa program studi yang cenderung kurang peminat, hal ini bisa disebabkan karena promosi yang dilakukan di UNISA Kuningan belum efektif dan efisien, padahal promosi merupakan hal penting dalam meningkatkan daya saing dan pemerataan mahasiswa bagi universitas. Oleh karena itu, pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru dalam strategi promosi. Menurut bagian PUSINFO, diketahui juga bahwa belum ada penelitian menggunakan data mahasiswa untuk mendapatkan informasi dalam melakukan strategi promosi. Pengolahan data mahasiswa dapat dilakukan dengan menggunakan *data mining*.

Data mining merupakan kegiatan dalam menggali suatu informasi yang tersembunyi dari *database* dengan melakukan penggalian pola-pola dari data yang hasilnya dapat digunakan sebagai pengambilan keputusan selanjutnya. Teknik *data mining* dalam mencari pola ataupun informasi baru salah satunya adalah *clustering*, yaitu proses untuk membagi data-data menjadi beberapa kelompok berdasarkan karakteristik yang telah

ditentukan. Banyak sekali algoritma yang dapat di gunakan untuk *clustering*, salah satunya Algoritma *K-Means clustering* dan *K-Medoids clustering*.

Data mining merupakan kegiatan dalam menggali suatu informasi yang tersembunyi dari *database* dengan melakukan penggalian pola-pola dari data yang hasilnya dapat digunakan sebagai pengambilan keputusan selanjutnya. Teknik *data mining* dalam mencari pola ataupun informasi baru salah satunya adalah *clustering*, yaitu proses untuk membagi data-data menjadi beberapa kelompok berdasarkan karakteristik yang telah ditentukan. Banyak sekali algoritma yang dapat di gunakan untuk *clustering*, salah satunya Algoritma *K-Means clustering* dan *K-Medoids clustering*.

K-Means adalah metode pengklasteran berbasis jarak dengan cara membagi data ke dalam beberapa *cluster* dimana setiap *cluster* memiliki tingkat variasi ketidaksamaan yang kecil. Sedangkan *K-Medoids* merupakan versi umum dari algoritma *K-Means* yang bekerja dengan mengukur jarak dan mempunyai komputasi yang lebih intensif. Keduanya sama-sama partisional (memecah *dataset* menjadi kelompok-kelompok) dan berusaha meminimalkan *squared error*, jarak antara titik berlabel dalam *cluster* dan titik yang ditunjuk sebagai pusat *cluster* (*centroid*). Pembeda antara *K-Medoids* dengan *K-Means* adalah *K-Medoids* memilih data *point* sebagai pusatnya.

Beberapa penulis terdahulu telah menerapkan teknik *K-Means dan K-Medoids* sebagai penelitian. Penelitian dengan menggunakan *K-Means Clustering* seperti yang dilakukan oleh (Yunita, 2018) dan (Widiyaningtyas, 2017), sedangkan penelitian dengan menggunakan *K-Medoids* dilakukan oleh (Dewi, 2017) yang dapat disimpulkan bahwa Algoritma *K-Means dan K-Medoids sama-sama* dapat digunakan dalam mengelompokkan data dengan efisien dan efektif dengan mendapatkan hasil yang diharapkan. Pada penelitian (Yunita, 2018) terdapat beberapa saran yang dapat dilakukan untuk penelitian selanjutnya yaitu dengan melakukan *data mining* dalam strategi promosi di universitas dengan menggunakan data pada setiap tahun ajaran baru dan dilakukan perbandingan dengan metode *clustering* lain. Penelitian (Nanda, 2016) melakukan perbandingan antara *K-Means* dan *K-Medoids* mempunyai saran untuk penelitian selanjutnya agar menggunakan dataset yang berbeda karena dengan dataset yang berbeda mungkin

memberikan hasil yang berbeda. Hal inilah yang mendasari penulis untuk melakukan penelitian dengan menggunakan perbandingan dari Algoritma *K-Means dan K-Medoids dari* data penerimaan mahasiswa baru tahun 2014 sampai tahun 2019 di UNISA Kuningan untuk mengetahui strategi promosi. Selain itu peneliti juga ingin mengetahui apakah terdapat perbedaan hasil dari kedua metode pengklusteran tersebut walaupun keduanya masih memiliki algoritma yang saling berkaitan sehingga mengetahui metode mana yang lebih baik untuk menentukan strategi promosi.

1.2 Rumusan Masalah

Rumusan masalah pada penulisan tesis ini adalah sebagai berikut:

1. Bagaimana mengetahui nilai k optimum menggunakan algoritma *K-Means* dan *K-Medoids* ?
2. Bagaimana hasil perbandingan pengolahan data menggunakan Algoritma *K-Means* dan *K-Medoids*?
3. Bagaimana strategi promosi yang tepat di UNISA Kuningan?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penulisan tesis ini adalah sebagai berikut :

1. Menentukan nilai k optimum untuk algoritma *K-Means* dan *K-Medoids* berdasarkan Dunn *Index*.
2. Mengetahui hasil perbandingan pengklasteran menggunakan metode *K-Means* dan *K-Medoids*.
3. Mengetahui strategi promosi yang tepat di UNISA Kuningan.

1.4 Batasan Masalah

Berdasarkan rumusan masalah yang telah diuraikan di atas, maka permasalahan dibatasi pada :

1 Data yang digunakan adalah data Mahasiswa Baru dari tahun 2014 sampai tahun 2019 di UNISA Kuningan.

2 *Software* yang digunakan untuk proses *data mining* adalah *Rapidminer 9.6*.

1.5 Sistematika Penulisan

Untuk memperoleh gambaran yang lebih jelas mengenai pembahasan mengenai penelitian ini, maka sistematika penulisannya sebagai berikut:

Bab I. Pendahuluan

Bab ini menjelaskan latar belakang yaitu alasan yang melatar belakangi penulis mengangkat permasalahan menjadi penelitian. Selain itu terdapat ruang lingkup yang menegaskan mengenai rumusan masalah, tujuan penelitan, dan batasan masalah.

Bab II. Landasan Teori

Bab ini menjelaskan tentang pengertian dan teori-teori yang digunakan sebagai penjelasan dan permasalahan yang dibahas. Teori terkait dijelaskan dari yang paling umum sampai yang palig khusus seperti promosi, *data mining*, *clustering*, *K-Means clustering, K-Medoids clustering,* dunn *Index*, dan aplikasi *rapidminer.*

Bab III. Objek dan Metodelogi Penelitian

Bab ini menjelaskan tentang gambaran umum tempat penelitan, sistem promosi berjalan,dan langkah-langkah yang digunakan dalam melakukan penelitian.

Bab IV. Hasil dan Pembahasan

Bab ini merupakan isi utama tesis yang menguraikan secara rinci mengenai pembahasan tentang pengelolaan *data mining* untuk menentukan strategi promosi.

Bab V. Kesimpulan dan Saran

Bab ini menjelaskan tentang kesimpulan penelitian serta saran-saran yang dapat diberikan untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI

2.1 Promosi

Promosi adalah salah satu bagian dari *marketing mix* yang mempunyai peran penting sebagai faktor penentu keberhasilan dalam pemasaran, karena seberapa berkualitasnya produk apabila konsumen belum mengetahui, mendengar dan belum yakin pada kegunaan produk tersebut, maka konsumen tidak akan membelinya. Sehingga dengan adanya promosi, maka perusahaan (penjual) dapat mendorong konsumen untuk membeli produk yang ditawarkan. Bauran promosi dikenal sebagai istilah dalam promosi, yang dapat di artikan sebagai kombinasi strategi yang paling baik dari variabel-variabel periklanan, *personal selling*, dan alat promosi lainnya agar dapat mencapai tujuan perusahaan dalam pemasaran produk (Kurniawati, 2017).

Menurut Kotler, peranan utama promosi adalah dengan mengadakan komuniksi kepada konsumen yang bersifat membujuk. Strategi promosi berdasarkan variabel-variabel yang ada pada *promotional mix*, yaitu *Advertising*, *Personal Selling, Sales Promotion, Public Relation, Direct Marketing* (Chasanah, 2017). Berikut adalah pengertian dari variabel yang ada pada *promotional mix* (Syarfan, 2016):

1. Periklanan (*Advertising*)

   Iklan adalah promosi dengan melakukan komunikasi secara tidak langsung dengan memberikan informasi mengenai keunggulan dari produk yang bertujuan untuk mengubah pikiran calon konsumen agar berminat untuk melakukan pembelian.

2. Penjualan Personal (*Personal Selling*).

   Personal selling adalah promosi dengan melakukan komunikasi secara langsung secara tatap muka antara penjual dan calon konsumen dengan tujuan untuk mengenalkan produk agar konsumen dapat memahami terhadap produk yang di tawarkan sehingga konsumen berminat untuk melakukan pembelian Promosi

3. Penjualan (*Sales Promotion*).

   Pemasaran pada penjualan adalah bentuk persuasi langsung kepada pelanggan dengan menggunakan berbagai insentif sehingga dapat meningkatkan jumlah pembelian produk.

4. Hubungan Masyarakat (*Public Relation*).

   Hubungan Masyarakat adalah promosi dengan melakukan komunikasi menyeluruh dari perusahaan dengan tujuan untuk dapat mempengaruhi keyakinan, opini, persepsi, serta sikap dari berbagai kelompok terhadap perusahaan.

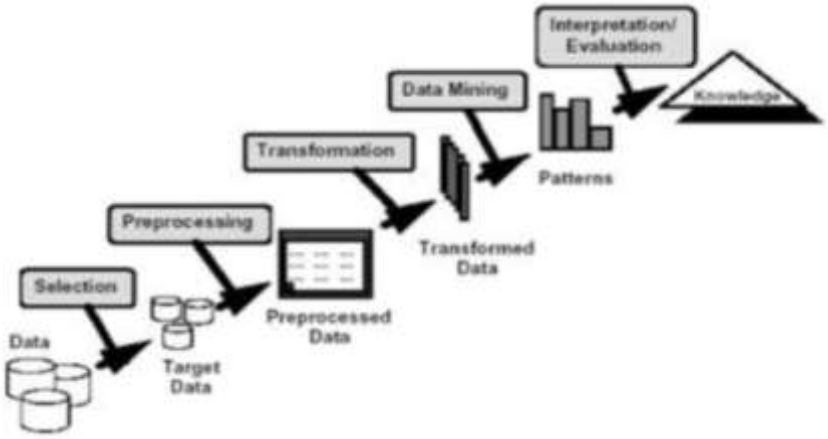
5. Pemasaran Langsung (*Direct Marketing*).

   Pemasaran secara langsung adalah promosi yang bersifat interaktif dengan cara memanfaatkan media iklan untuk mendapatkan respon dari konsumen. Transaksi untuk pemasaran ini dilakukan disembarang lokasi.

2.2 *Data Mining*

Data mining dilakukan dengan cara mengekstraksi dan mengenali pola yang menarik pada *database* sehingga menghasilkan informasi yang jika secara manual informasi tersebut belum dapat diketahui (Vulandari, 2017). *Data mining* juga merupakan suatu proses penyelesaian masalah dengan cara menganalisis data yang terdapat pada *database* dimana proses pencariandilakukan secara otomatis seperti pada komputer (Maimon, 2010). Dari beberapa pengertian *data mining* tersebut dapat ditarik kesimpulan bahwa *data mining* merupakan kegiatan penggalian informasi yang tersembunyi pada *database* dengan cara menganalisis data menggunakan teknik-teknik tertentu, yang kemudian hasilnya berupa pola-pola yang dapat digunakan sebagai pengambilan keputusan di masa yang akan datang.

Data mining merupakan salah satu rangkaian pada *Knowledge Discovery in Databases* (KDD). Gambar 2.1 menjelaskan rangkaian proses yang terjadi dalam *Knowledge Discovery in Databases* (KDD) :



Gambar 2.1
Tahapan KDD (Pramesti, 2017)

Gambar 2.1 menjelaskan tahapan-tahapan dari KDD, yaitu:

1. *Selection* yaitu proses pemilihan data yang relevan dari *database* agar penganalisisan dapat di lakukan.
2. *Preprocessing,* yaitu proses untuk membersihkan data dengan cara membersihkan data yang tidak relevan dan proses untuk mengintegrasikan data dengan cara menggabungkan data dari banyak sumber.
3. *Transformation*, dan data yang ditransformasikan ke dalam bentuk yang sesuai agar dapat dilakukan proses penambangan dengan melakukan agregasi (operasi ringkasan).
4. *Data mining*, yaitu proses melakukan penggalian data dengan menggunakan algoritma,teknik, atau, metode tertentu untuk melakukan ekstraksi pola data.
5. *Pattern Evaluation* yaitu proses mengevaluasi pola dengan cara mengidentifiksi pola-pola yang dengan jelas mempresentasikan pengetahuan berdasarkan langkah-langkah sebelumnya.
6. *Interpretation Knowledge*, proses untuk mempersentasikan pengetahuan dengan menggunakan teknik visualisasi dan representasi pengetahuan yang bertujuan untuk menyajikan pengetahuan yang telah ditambang kepada pengguna agar mudah untuk dipahami.

2.2.1 *Preprocessing* Data

Preprocessing data harus dilakukan karena dalam *database* biasanya banyak data mentah yang tidak dapat diproses (Larose, 2014), seperti :

1. Bidang yang sudah usang (kadaluwarsa).
2. Nilai yang hilang (*missing values*).
3. Pencilan (*outlier).*
4. Data dalam bentuk yang tidak cocok untuk model penambangan data, contohnya data yang masih berbentuk tumpukan kertas.
5. Nilai tidak konsisten dengan kebijakan atau akal sehat.

Pembersihan data berfungsi untuk memperbaiki data yang tidak sesuai seperti data yang kosong, *outlier*, dan mengurangi data yang tidak perlu. Dalam menangani nilai-nilai yang hilang biasanya dilakukan dengan menghilangkan *record* data. Metode ini sangat disayangkan jika nilai yang hilang hanya satu, karena akan menghilangkan informasi di semua bidang lain, hanya karena satu nilai bidang yang tidak ada. Kriteria umum untuk mengganti data yang hilang adalah sebagai berikut (Larose, 2014):

1. Ganti nilai yang hilang dengan konstanta, yang ditentukan oleh analis. Contohnya pada Gambar 2.2.

Gambar 2.2
Merubah data kosong dengan konstanta yang ditentukan analis.

Pada Gambar 2.2, data kosong pada *record* dua "*cubicinches*" diisi dengan angka nol dan *record* 4 pada *brand* diisi dengan "*Missing*".

2. Ganti nilai yang hilang dengan bidang mean (untuk variabel numerik) atau mode/modus (untuk variabel kategori).

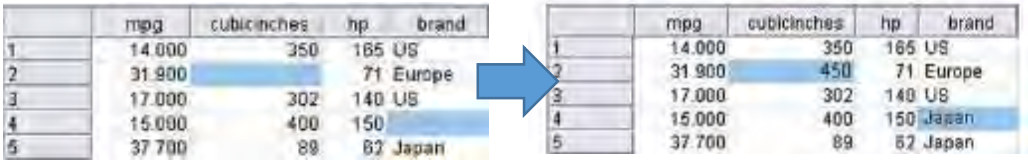
	mpg	cubicinches	hp	brand
1	14.000	350	165	US
2	31.900		71	Europe
3	17.000	302	140	US
4	15.000	400	150	
5	37.700	89	62	Japan

	mpg	cubicinches	hp	brand
1	14.000	350	165	US
2	31.900	200.65	71	Europe
3	17.000	302	140	US
4	15.000	400	150	US
5	37.700	89	62	Japan

Gambar 2.3
Merubah data kosong dengan *mean* dan modus

Pada Gambar 2.3, data kosong pada *record* dua "*cubicinches*" diisi dengan rata-rata data pada kolom "cubicinches" dan *record* 4 pada *brand* diisi dengan nilai yang paling banyak keluar.

3. Ganti nilai yang hilang dengan nilai yang dihasilkan secara acak/random dari yang diamati distribusi variabel.



	mpg	cubicinches	hp	brand
1	14.000	350	165	US
2	31.900		71	Europe
3	17.000	302	140	US
4	15.000	400	150	
5	37.700	89	62	Japan

	mpg	cubicinches	hp	brand
1	14.000	350	165	US
2	31.900	450	71	Europe
3	17.000	302	140	US
4	15.000	400	150	Japan
5	37.700	89	62	Japan

Gambar 2.4
Merubah data kosong secara acak

Pada Gambar 2.4, data kosong pada *record* dua "*cubicinches*" dan *record* 4 pada *brand* diisi dengan nilai yang dihasilkan secara acak dari data pada kolomnya masing-masing.

2.2.2 Clustering

Clustering merupakan proses untuk mengelompokkan data (objek) agar objek-objek yang mirip (berhubungan) satu sama lain berada dalam satu *cluster* dan objek-objek yang berbeda (tidak berhubungan) berada dalam *cluster* lain (Talakua, 2017). Suatu *cluster* dapat dikatakan baik apabila *cluster* mempunyai:

1. Homogenitas (kesamaan), yaitu setiap data dalam satu *cluster* memiliki tingkat kesamaan yang tinggi. Contohnya pada *cluster* mahasiswa yang mengutamakan nilai akhir yang bagus akan terdiri dari mahasiswa yang mengutamakan belajar dan melakukan setiap kegiatan perkuliahan dengan sungguh-sungguh, sedangkan mahasiswa yang lebih sering tidak masuk kuliah tentu tidak dapat digabungkan menjadi anggota *cluster*.

2. Heterogenitas (perbedaan), yaitu antara satu *cluster* dengan *cluster* lainnya memiliki tingkat perbedaan yang tinggi. Dalam contoh pertama, anggota dari *cluster* mahasiswa yang mengutamakan nilai akhir yang bagus tentu mempunyai pendapat yang jelas berbeda dengan anggota-anggota *cluster* mahasiswa yang sering tidak masuk kuliah.

Salah satu metode dalam proses *clustering* adalah metode non hirarki (*partitional clustering)*. Pada metode ini, setiap *cluster* memiliki titik pusat *cluster* (*centroid*), dimana setiap *cluster* terdiri dari data yang tidak mempunyai struktur hirarki antara satu dengan yang lainnya. Tujuan metode ini adalah untuk meminimumkan jarak (*dissimilarity*) dari seluruh data ke *centroid* masing-masing. Untuk memulai *partitional clustering*, jumlah *cluster* harus di tentukan terlebih dahulu sesuai yang diinginkan (Irwansyah, 2017). Teknik *partitional clustering* yaitu membagi objek menjadi beberapa partisi di mana satu partisi menggambarkan *cluster*, dimana objek dalam satu *cluster* memiliki karakteristik yang serupa dan objek pada *cluster* yang lain memiliki karakteristik yang berbeda. *K-mean* dan *K-Medoids* adalah contoh dari *partitional clustering* (Gandhi, 2014).

Algoritma *clustering* yang berbeda akan menunjukkan hasil yang berbeda, karena sangat sensitif terhadap karakteristik kumpulan data asli khususnya *noise* dan *dimension*. Kualitas proses pengelompokan tersebut menentukan kemurnian *cluster* dan karenanya sangat penting untuk mengevaluasi hasil dari algoritma *clustering*. Sehingga, kegiatan validasi *clustering* telah menjadi tugas utama. Faktor utama yang mempengaruhi validasi *clustering* adalah indeks validitas untuk memilih *Number of Cluster* (NC). Indeks validitas digunakan untuk mengukur kualitas hasil *clustering*. Ada dua jenis indeks validitas, yaitu indeks eksternal dan indeks internal. Indeks eksternal adalah ukuran perjanjian antara dua partisi di mana partisi pertama adalah struktur pengelompokan apriori, dan hasil kedua dari prosedur. Indeks internal digunakan untuk mengukur kebaikan struktur pengelompokan tanpa menggunakan informasi eksternal. Untuk indeks eksternal, (Wang, 2009) mengevaluasi hasil dari algoritma pengelompokan berdasarkan struktur *cluster* yang diketahui dari kumpulan data (label *cluster*), sedangkan indeks internal mengevaluasi

hasilnya menggunakan jumlah dan fitur yang melekat dalam kumpulan data. NC yang optimal biasanya ditentukan berdasarkan internal indeks.

Prosedur umum untuk menentukan partisi terbaik dan jumlah *cluster* optimal dari satu set objek dengan menggunakan langkah-langkah validasi internal adalah sebagai berikut.

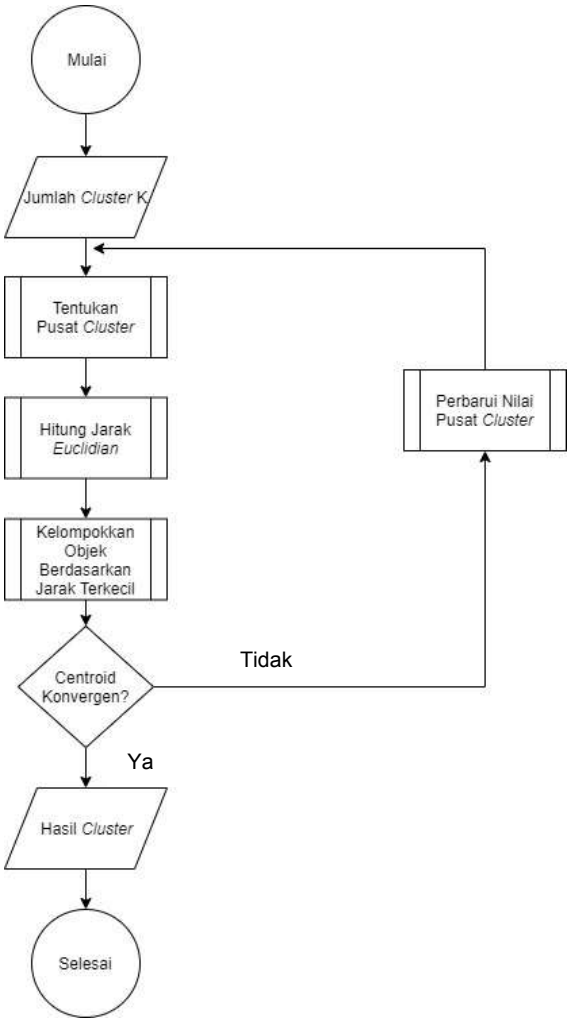
1. Inisialisasi daftar algoritma pengelompokan yang akan diterapkan pada sekumpulan data.
2. Untuk setiap algoritma pengelompokan, gunakan kombinasi parameter yang berbeda untuk mendapatkan hasil pengelompokan yang berbeda.
3. Hitung indeks validasi internal yang sesuai dari setiap partisi yang diperoleh pada Langkah 2.
4. Pilih partisi terbaik dan jumlah *cluster* optimal sesuai dengan kriteria.

2.2.3 Algoritma *K-Means*

Algoritma *K-Means* merupakan salah satu algoritma pada *partitional clustering* yang mempunyai tujuan untuk mengelompokan data, dimana setiap kelompok yang memiliki kesamaan pada karakteristiknya dikelompokan dalam satu kelompok dan data yang karakteristiknya berbeda dikelompokan pada kelompok lain (Sibarani, 2018). Algoritma ini merupakan algoritma yang paling banyak digunakan karena mempunyai kemampuan dalam mengelompokkan data dengan waktu komputasi yang cepat dan efisien dengan menggunakan jumlah data yang cukup besar. Ide dasar algoritma *K-Means* sangatlah sederhana, yaitu meminimalkan *Sum of Squared Error* (SSE) antara objek-objek data dengan sejumlah k pada *centroid* (Irwansyah, 2017).

K-Means merupakan algoritma *clustering* yang berulang-ulang dengan menetapkan jumlah dari *cluster* (k) secara acak, dimana nilai tersebut menjadi pusat dari *cluster* untuk sementara. Pusat *cluster* biasa disebut dengan *centroid*, *mean* atau *means.* (Vulandari, 2017). *K-Means* merupakan algoritma yang mudah untuk diimplementasikan, cepat, dan dapat bekerja pada berbagai jenis data dimana hasil pengelompokannya bergantung pada *centroid*, karena jika *centroid* nya tidak sesuai maka hasil pengelompokan akan lebih tidak

stabil dan jumlah iterasi akan meningkat (Guoli, 2013). Dari definisi Algoritma *K-Means* sebelumnya dapat ditarik kesimpulan bahwa *K-Mean clustering* merupakan metode *clustering non-hirarki* yang mudah untuk di implementasikan dan bergantung pada *centroid* yang berasal dari penentuan jumlah *cluster* k secara random agar pengelompokkan data sesuai dengan karakteristiknya sehingga pada setiap kelompoknya memiliki tingkat ketidaksamaan yang kecil. Langkah-langkah *K-Means Clustering* dijelaskan pada Gambar 2.5:



Gambar 2.5
Flowchart Algoritma *K-Means*

Berdasarkan Gambar 2.5 dapat dijelaskan langkah-langkah Algoritma *K-Means* sebagai berikut:

1. Dari himpunan data yang akan di klasterisasi, di pilih sejumlah k objek secara acak.

2. Dari jumlah k objek, tentukan *centroid*.
3. Hitung jarak antar objek dengan *centroid* menggunakan rumus *Euclidean Distance* yang dirumuskan sebagi berikut (Nishom, 2019) :

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2.1)$$

dimana:

d = jarak antara x dan y

x = data pusat *cluster*

y = data pada atribut

i = setiap data

n = jumlah data

x_i = data pada pusat *cluster* i

y_i = data pada setiap data ke i

4. Kelompokkan setiap objek berdasarkan jarak terkecil.
5. Lihat hasil akhir *cluster,* jika belum stabil (*konvergen*) maka perbarui nilai *centroid* lama dengan nilai *centroid* baru dan lakukan kembali proses pengulangan (iterasi) perhitungan jarak hingga semua *centroid* stabil (konvergen). *Konvergen* artinya hasil iterasi pada semua *centroid* paling akhir sama dengan nilai semua *centroid* pada iterasi sebelumnya.

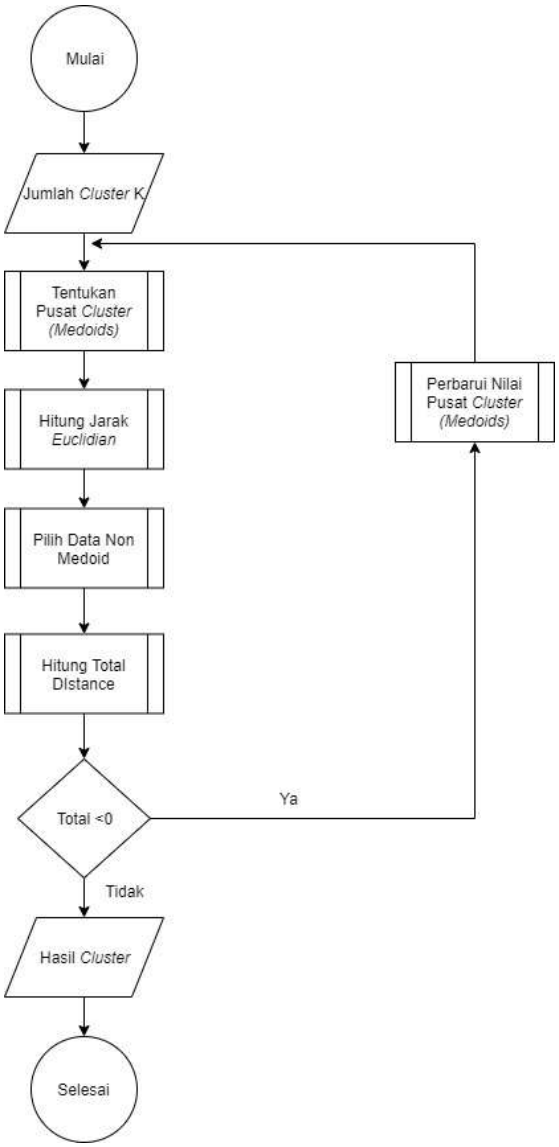
2.2.4 Algoritma *K-Medoids*

K-Medoids Clustering dikenal juga sebagai *Partitioning Around Medoids* (*PAM*), dimana penggunaan medoid bukan dari pengamatan *mean* yang dimiliki oleh setiap *cluster*, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset (Vercellis, 2009).

Algoritma *K-Medoids* menggunakan teknik berbasis objek *representative* (perwakilan) yang mempunyai keunggulan untuk menghadapi *noise* dan *outlier* dengan cara menghilangkan penggunaan rata-rata, sehingga dapat memperbarui *centroid* dan menggantinya dengan objek aktual sebagai representasi dari suatu *cluster*. Algoritma *K-*

Medoids melakukan partisi dengan cara meminimalkan jumlah *dissimilarity* antara setiap objek p dan objek *representative* terdekat dengan menggunakan jumlah kesalahan absolut (Suyanto, 2017).

Langkah-langkah *K-Medoids Clustering* di jelaskan pada Gambarl 2.3 :



Gambar 2.6
Flowchart Algoritma *K-Medoids*

Berdasarkan Gambar 2.6 dapat dijelaskan langkah-langkah Algoritma *K-Medoids* sebagai berikut:

1. Pilih sejumlah k objek secara acak.
2. Dari jumlah k objek, tentukan *centroid* perwalian (medoid).

3. Masukan setiap objek ke dalam *cluster* terdekat berdasarkan jarak menggunakan rumus *Euclidean Distance*.
4. Pilih secara acak sebuah objek non medoid.
5. Hitung total *distance* (S) dengan menghitung nilai TD baru - TD lama. Jika *S* < 0, tukar objek dengan data *cluster* untuk membentuk sekumpulan *k* objek baru sebagai *medoid* (Pramesti, 2017). Lakukan proses iterasi perhitungan jarak dan total *distance* (S) sehingga tidak terjadi perubahan medoid.

2.3 Dunn *Index*

Dunn indeks merupakan salah satu indeks validasi internal, dimana indeks validasi bertujuan untuk mengevaluasi hasil pada *clustering* sehingga menemukan jumlah *cluster* terbaik dan dihitung berdasarkan (Liu, 2010) *compactness dan separation. Compactness* merupakan tingkat kesamaan objek dalam *cluster* yang sama, sedangkan *separation* merupakan tingkat perbedaan objek dalam *cluster* yang berbeda. Terdapat validasi internal dan validasi eksternal, dimana validasi internal mengandalkan informasi di dalam data, sedangkan validasi eksternal mengandalkan informasi di luar data.

Proses perhitungan Dunn Indeks adalah dengan menghitung nilai minimum dari perbandingan antara nilai fungsi dissimilaritas antara dua *cluster* sebagai *separation* dan nilai maksimum dari diameter *cluster* sebagai *compactness*. Jumlah *cluster* terbaik ditunjukkan dengan semakin besarnya nilai Dunn. Berikut adalah rumus Dunn Indeks (Chowdary, 2014):

$$D = \min_{1 \le i \le n} \left(\min_{1 \le j \le n, i \ne j} \left(\frac{d(i,j)}{max_{1 \le k \le n} d(k)} \right) \right) \quad (2.2)$$

Dimana :

$d(i,j)$ = jarak antara *cluster* i dan j,

$d(k)$ = mengukur jarak intra*cluster cluster* k.

2.4 *Rapidminer*

RapidMiner merupakan perangakat lunak yang bersifat terbuka (*open source*) yang dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari University of Dortmund dengan munggunakan bahasa java dan dapat berjalan di semua sistem operasi. Rapidminer merupakan solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi, karena dalam memberikan wawasan kepada pengguna *rapideminer* menggunakan teknik deskriptif dan prediksi.

Rapidminer dapat membuat keputusan yang paling baik karena memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *preprocessing data* dan visualisasi. *Rapidminer* dapat digunakan untuk mendukung langkah-langkah dalam proses pembelajaran mesin termasuk untuk persiapan data, hasil visualisasi, validasi model, dan optimasi.Selain itu, *rapidminer* juga dapat digunakan untuk bisnis, penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi. (Klinkenberg, 2013; Sibuea, 2017).

2.5 Penelitian Terkait

Berikut ini merupakan beberapa penelitian yang terkait dengan *clustering* menggunakan *K-Means* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Penelitian Terkait

No	Peneliti	Hasil
1	(Yunita, 2018)	Peneltian ini bertujuan untuk melakukan *clustering* teradap penerimaan mahasiswa baru menggunakan algoritma *K-Means*. Penelitian ini menghasilkan tiga *cluster* dengan tiga kali iterasi. Menurut penelitian ini, titik pusta *K-Means* sangat berpengaruh pada hasil akhir *cluster*, dan strategi promosi akan mengikuti *cluster* yang terbentuk bardasarkan program studi yang paling banyak diminati di masing-masing sekolah. Penelitian ini diharapkan dapat dijadikan sebagai referensi untuk penelitian selanjunya dan melakukan perbandingan dengan metode *clustering* lain.
2	(Monalisa, 2018)	Penelitian ini melakukan pengelompokkan pelanggan menggunakan model LRFM dengan menggunakan validasi dunn *index* dan silhoute *index,* dimana kedua validasi ini mengambil nilai paling tinggi sebagai K terbaik. Hasil akhir nya K terbaik ada pada *cluster* 3 dimana nilai Dunn Index 0,84 dan nilai silhoute *index*

No	Peneliti	Hasil
		0,54. Sehingga kedua validasi tersebut dapat digunakan untuk mengelompokkan pelanggan.
3	(Marlina, 2018)	Pada penelitian ini melakukan perbandingan antara algoritma *K-Means* dan *K-Medoids* dalam pengelompokkan wilayah sebaran cacat pada anak menggunakan validasi *silhoutte index* dimana hasil akhirnya *K-Medoids* menjadi algoritma paling baik dengan nilai *silhoutte* 0,5009 sedangkan *K-Means* 0,1443.
4	(Widiyaningtyas, 2017)	Penelitian ini menggunakan algoritma *K-Means clustering* yang untuk menganalisis distribusi guru pada sekolah menengah di Indonesia. Penelitian ini terbagi menjadi dua belas *cluster* dengan nilai *Sum of Squared Error* (SSE) dengan persentase 87,15%. Berdasarkan hasil yang diperoleh keakuratan Algoritma *K-Means clustering* dengan dua belas *cluster* sangat tinggi.
5	(Dewi, 2017)	Penelitian ini menggunakan algoritma *K-Medoids* untuk mengetahui pola produksi suatu industri agar dapat memberikan keputusan untuk pengembangan dan kemajuan industri. Pengelompokan ini terbagi menjadi dua *cluster*, dengan tiga kali iterasi. Hasil pengelompokan memiliki kualitas kemurnian senilai satu yang berarti kualitas *cluster* pada *K-Medoids* baik.
6	(Hermanto, 2017)	Pada penelitian ini dilakukan perbandingan *K-Means* dan *K-Medoids* dengan validasi *Davies Bouldin Index* untuk menganalisa algoritma yang paling baik dalam mengelompokkan jenis bunga dengan menggunakan dataset Iris. Hasil akhirnya algoritma *K-Medoids* memiliki nilai *Davies Bouldin* sebesar 0.291 dan algoritma *K-Means* memiliki nilai *Davies Bouldin* paling rendah yaitu 0.167. Sehingga dalam hal ini algoritma yang dapat digunakan untuk menentukan jenis dari bunga iris sendiri dapat menggunakan algoritma *K-Means* untuk mempermudah dalam proses *clustering*.
7	(Nanda, 2016)	Penelitian melakukan perbandingan antara *K-Means* dan *K-Medoids* menggunakan dataset secara acak. Berdasarkan kumpulan data ini, *K-Medoids* menjadi algoritma terbaik karena sesuai analisis dan menghabiskan lebih sedikit waktu. Untuk peneltian berikutnya, menggunakan dataset yang berbeda untuk menguji efisiensi pada kedua algoritma, karena dengan dataset yang berbeda mungkin memberikan hasil yang berbeda.
8	(Rendón, 2011)	Penelitian ini melakukan perbandingan antara empat validasi internal dan empat validasi eksternal dengan menggunakan 13 dataset yang dikelompokkan dengan menggunakan Algoritma *K-Means* dan *Bissection K-Means* dengan hasil akhir nilai kebenaran validasi internal 86%, dan 51,9% untuk validasi eksternal.

Berdasarkan Tabel 2.1, pada penelitian (Yunita, 2018) yang mempunyai beberapa saran untuk penelitian selanjutnya yaitu dengan melakukan *data mining* dengan data pada setiap tahun ajaran baru dan dilakukan perbandingan dengan metode *clustering* lain. Selain itu juga menurutnya bahwa titik pusat k sangat berpengaruh pada hasil akhir *cluster*. Selain itu, penelitian (Nanda, 2016) melakukan perbandingan antara *K-Means* dan *K-Medoids* terdapat saran untuk penelitian selanjutnya agar menggunakan dataset yang berbeda karena dengan dataset yang berbeda mungkin memberikan hasil yang berbeda.

Dengan demikian pada penelitian ini akan dilakukan proses *data mining* dengan melakukan perbandingan antara *K-Means* dan *K-Medoids*. Selanjutnya, dalam menentukan *centroid* k terbaik, penelitian ini menggunakan validasi internal karena pada penelitian (Rendón, 2011)*,* tingkat kebenaran validasi internal lebih besar dari validasi eksternal. Validasi internal yang digunakan pada penelitian ini adalah Dunn Indeks karena berdasarkan Tabel 2.2 belum ada penelitian yang membandingkan Algoritma *K-Means* dan Algoritma *K-Medoids* menggunakan Dunn Indeks.

Tabel 2.2 Perbandingan Penelitian

Penelitian	***K-Means***	***K-Medoids***	**Algoritma Lain**	**Validasi**
Yunita, 2018	√	-	-	-
Monalisa, 2018	√	-	-	*Dunn Index dan Silhoutte Index*
Marlina, 2018	√	√	-	*Silhoutte Index*
Widiyaningtyas, 2017	√	-	-	-
Dewi, 2017	-	√	-	-
Hermanto, 2017	√	√	-	*Davies Bouldin Index*
Nanda, 2016	√	√	-	-
Rendón, 2011	√	-	*Bissection K-Means*	*BIC, Calinski-Harabasz, Davies-Bouldin, Silhoutte, Dunn, F-measure , Nmimeasure, Purity, Entropy*
Neng Sri, 2019	√	√	-	*Dunn Index*

BAB III

OBJEK DAN METODOLOGI PENELITIAN

3.1 Profil UNISA Kuningan

Penelitian berorientasi pada strategi promosi di Universitas Islam Al-Ihya (UNISA) Kuningan yang merupakan perguruan tinggi swasta yang terletak di Jl. Mayasih No 11 Cigugur Kota/Kabupaten Kuningan Provinsi Jawa Barat yang telah resmi dan diakui oleh Kementrian Pendidikan dan Kebudayaan Republik Indonesia. Pada tanggal 18 April 1986, Yayasan *Islamic Centre* Al-Ihya dengan didukung Pemerintah Daerah Kabupaten Kuningan prakarsa dalam mendirikan Fakultas Tarbiyah Perguruan Tinggi Agama Islam (PTI) Al Ihya. Tahun 1987 dengan Keputusan Kordinator Kopertais Wilayah II Jawa Barat nomor : 05 tahun 1987 diperoleh ijin operasionalnya. Status ijin operasional berjalan dari tahun 1987 sampai tahun 1989. Sejak itu status dari Fakultas Tarbiyah PTI Al Ihya Kuningan berubah nama menjadi Sekolah Tinggi Tarbiyah Islamiyah (STTI) Al-Ihya dengan status meningkat menjadi "TERDAFTAR" berdasarkan Keputusan Menteri Agama Nomor : 114 tahun 1989.

Pada tahun 1993 STTI Al-Ihya dikembangkan menjadi Sekolah Tinggi Agama Islam (STAI) Al-Ihya Kuningan sampai tahun 2014. Pada tanggal 16 Mei 2014, STAI Al-Ihya Kuningan berubah menjadi Universitas Islam Al-Ihya (UNISA) Kuningan dan menjadi perguruan tinggi resmi yang diakui pemerintah sesuai dengan SK PT : 95/E/O/2014. UNISA Kuningan memiliki empat fakultas dengan tiga belas program studi. Berikut adalah Visi dan Misi UNISA Kuningan :

VISI

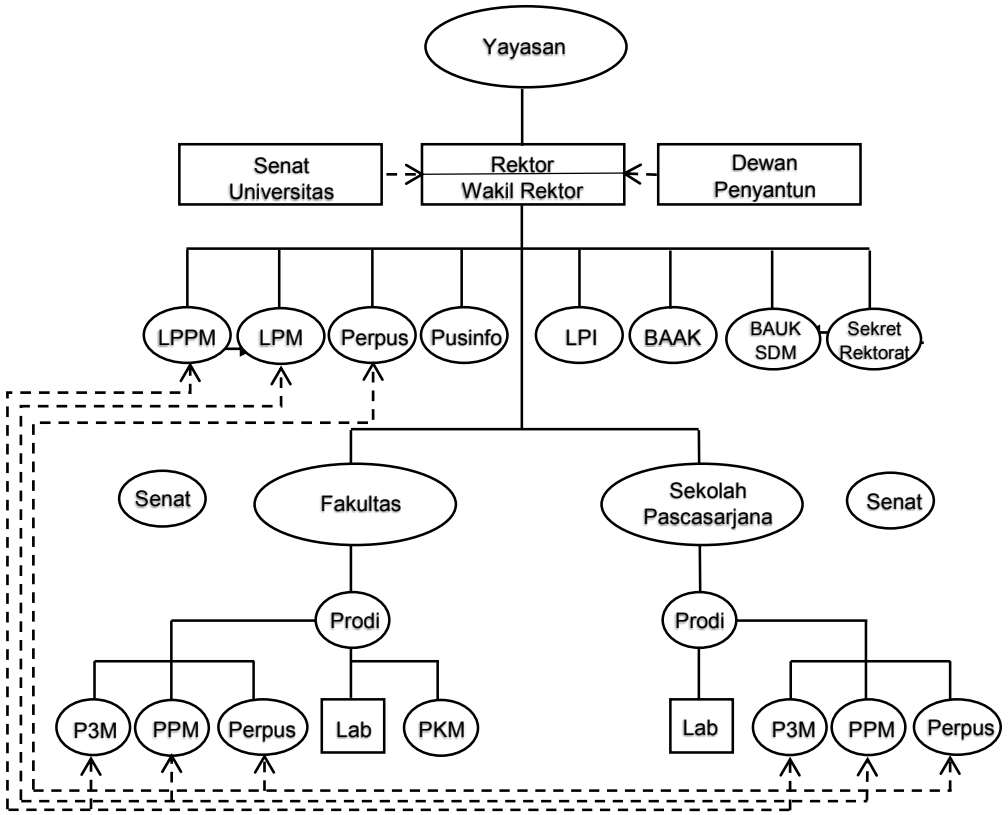
"Menjadi Perguruan Tinggi Islam Terdepan di Bidang Kewirusahaan Berbasis *Information Communication Technology* Tahun 2026".

MISI

1. Menghasilkan lulusan yang memiliki jiwa kewirausahaan, menguasai teknologi informasi dan komunikasi, mandiri, professional, kompetetif dan berakhlak mulia.

2. Mencetak ilmuwan muda yang memiliki integritas, berdedikasi tinggi serta memiliki komitmen untuk mengembangkan ilmu pengetahuan dan teknologi.
3. Menjadi pusat pengembangan kewirausahaan, teknologi dan seni serta menggunakan pengupayaannya untuk meningkatkan taraf kehidupan masyarakat.
4. Menjadi mitra pemerintah dalam pelaksanaan pembangunan masyarakat di segala bidang kehidupan
5. Menjadi bagian dari masyarakat dalam upaya mengembangkan hidup dan kehidupan yang lebih baik serta berakhlakul karimah.
6. Menjadi pelayan dan pendamping masyarakat untuk tumbuh menjadi masyarakat yang agamis, cerdas dikir, cerdas fikir, kreatif dan mandiri.

Struktur organisasi Universitas Al-Ihya Kuningan seperti pada Gambar 3.1 :



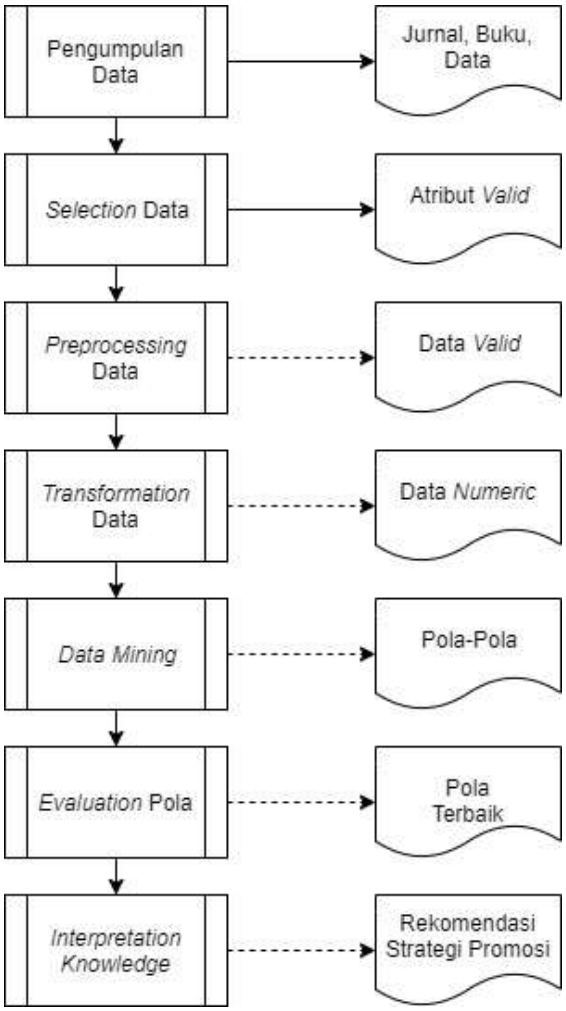
Keterangan :

LPPM	: Lembaga Penelitian dan Pengabdian Masyarakat
LPI	: Lembaga Pengawasan Internal
LPM	: Lembaga Penjaminan Mutu
BAAK	: Biro Administrasi Akademik dan Kemahasiswaan
BAUKSDM	: Biro Administrasi Umum, Keuangan, Sarana Prasarana, dan Sumber Daya Manusia
P3M	: Pusat Penelitian dan Pengabdian Masyarakat
PPM	: Pusat Penjaminan Mutu
PKM	: Pusat Kendali Mutu

Gambar 3.1
Struktur Organisasi UNISA

3.2 Metodologi Penelitian

Gambar 3.2 merupakan tahapan-tahapan penelitian yang dilakukan pada penelitian ini.



Gambar 3.2
Metodologi Penelitian

Dari Gambar 3.2, dapat di uraikan langkah-langkah analisis penelitian sebagai berkut:

1. Pengumpulan Data

   Pengumpulan data didapatkan berdasarkan sumber data yang dimilki. Sumber data tersebut ada dua, yaitu:

   a. Studi Lapangan

      Studi lapangan dilakukan dengan cara melakukan observasi, wawancara, dan mendapatkan data calon mahasiswa. Observasi digunakan untuk melihat

keadaan nyata di UNISA yang menjadi objek penelitian, wawancara untuk mengetahui strategi promosi apa saja yang pernah dilakukan, dan memperoleh data mahasiswa sebagai input data yang akan diolah.

b. Studi Pustaka

Studi pustaka dilakukan dengan cara mencari dan menelaah buku, jurnal, ataupun penelitian terdahulu yang menunjang topik yang akan diteliti. Hal ini bertujuan agar penelitian memiliki dasar pengetahuan yang sesuai dengan arah penelitian yang akan dilakukan.

2. *Selection Data*

Tidak semua data pada *database* dapat digunakan, hanya data yang sesuai atau berpengaruh untuk di lakukan analisis sebagai strategi promosi saja yang akan diambil dari *database*. Contoh data yang tidak berpengaruh pada strategi promosi seperti nama calon mahasiswa karena bersifat unik.

3. *Preprocessing Data*

Penelitian ini melakukan *prepocessing data* berdasarkan proses K*nowledge Discovery in Database* (KDD). Pembersihan data dan Integrasi data akan dilakukan pada penelitian ini dengan mengganti data kosong dengan *mean*, *median*, *modus*, ataupun menghilangkan data yang mempunyai banyak data kosong sehingga akan menghasilkan data yang valid.

4. *Data Transformation*

Tahap ini merupakan tahap untuk mengubah bentuk data seperti data pada variabel alamat menjadi kecamatan dan semua data yang bertipe teks akan diubah kedalam bentuk numerik.

5. *Data Mining*

K-Means Clustering dan *K-Medoids Clustering* merupakan algoritma yang akan digunakan pada penelitian ini dengan menggunakan aplikasi *Rapidminer* sehingga akan menghasilkan pola-pola yang unik sesuai dengan jumlah K yang dilakukan percobaan.

6. *Pattern Evaluation*

   Pada tahapan ini, hasil dari pola-pola unik akan dilakukan pengevaluasian untuk mengetahui pola mana yang terbaik dan algoritma mana yang paling baik. Untuk mengetahui pola dan algoritma terbaik, peneliti menggunakan validasi Dunn Indeks.

7. *Interpretation Knowledge*

   Interpretation Knowledge adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis strategi promosi yang didapat dari tabel perbandingan *K-Means dan K-Medoids* dengan validasi Dunn Indeks berupa teks dan tabel.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data penelitian diperoleh data mentah mahasiswa baru dari bagian PUSINFO di UNISA Kuningan dari tahun 2014 sampai tahun 2019 dalam bentuk format *excel*. Data mahasiswa baru terpisah untuk setiap tahunnya dengan total *record* dari tahun 2014 sampai tahun 2019 adalah 2.047 *record*. Ada 16 atribut pada data mahasiswa baru UNISA Kuningan seperti pada Tabel 4.1 :

Tabel 4.1 Daftar Atribut Data

No	Atribut	Type Atribut	Keterangan
1.	No	*Integer*	Nomor urut
2.	NPM	*Real*	Nomor Pokok Mahasiswa
3.	Nama Mahasiswa	*Polynominal*	Nama Mahasiswa
4.	L/P	*Polynominal*	Jenis Kelamin L= Laki-Laki P= Perempuan
5.	Prodi	*Polynominal*	Program Studi 1. Ilmu Gizi 2. Kesehatan Masyarakat 3. Komunikasi Penyiaran Islam (KPI) 4. Pendidikan Agama Islam (PAI) 5. Pendidikan Bahasa Inggris (PBO) 6. Pendidikan Guru Madrsah Ibtidaiyah (PGMI) 7. Pendidikan Guru Pendidikan Anak Usia Dini (PGPAUD) 8. Pendidikan Guru Sekolah Dasar (PGSD) 9. Pendidikan Jasmani, Kesehatan dan Rekreasi (PJKR) 10. Perbankan Syari'ah (PS) 11. Teknik Informatika (TI) 12. Teknik Mesin (TM) 13. Teknologi Pangan (TP)
6.	Tempat Lahir	*Polynominal*	Tempat Lahir Mahasiswa
7.	Tanggal Lahir	*Date*	Tanggal Lahir Mahasiswa
8.	Umur	*Integer*	Umur Mahasiswa
9.	Asal Sekolah	*Polynominal*	Asal Sekolah Mahasiswa
10.	Program	*Polynominal*	Jurusan Asal Sekolah
11.	Status Sekolah	*Polynominal*	Status Asal Sekolah
12.	Nilai UN	*Real*	Nilai Ujian Nasional
13.	Kelas	*Polynominal*	Kelas Mahasiswa 1. Karyawan 2. Reguler

No	Atribut	Type Atribut	Keterangan
14.	Media	*Polynominal*	Media Promosi Universitas 1. Brosur 2. Internet 3. Rekomendasi 4. Spanduk
15.	Alamat	*Polynominal*	Alamat Mahasiswa
16.	Gaji Orang Tua	*Currency*	Gaji Orang Tua 1. Rp. 1.000.000 s/d Rp. 1.900.000 2. Rp. 2.000.000 s/d Rp. 2.900.000 3. Rp. 3.000.000 s/d Rp. 3.900.000 4. Rp. 4.000.000 s/d Rp. 4.900.000 5. Rp. 5.000.000 s/d Rp. 5.900.000 6. Rp. 6.000.000 s/d Rp. 6.900.000

Tabel 4.2 merupakan data mentah mahasiswa baru pada Tahun 2014 dengan jumlah 324 *record*.

Tabel 4.2 Data Mahasiswa Baru UNISA Kuningan Tahun 2019

No	NPM	Nama Mahasiswa	L/P	Prodi	Tempat Lahir	Tanggal Lahir	Umur	Asal Sekolah	Program	Status Sekolah	Nilai UN	KELAS	Media	Alamat	Gaji Orang Tua (Rp.)
1	14152 21002	Ika Kartika	P	Ilmu Gizi	Kuningan	03/ 01/ 1995	19	SMAN 1 Cigugur	IPA	Negeri	45,32	KARYA WAN	Brosur	Kuningan Kab.Kuningan	2.000.000 s/d 2.900.000
2	14152 25018	Maman Rohman	L	Ilmu Gizi	Kuningan	03/ 01/ 1995	19	SMK YAMSIK Kuningan	RPL	Swasta	60,25	KARYA WAN	Brosur	Kuningan Kab.Kuningan	3.000.000 s/d 3.900.000
3	14152 23009	Nuraeni	P	Ilmu Gizi	Kuningan	01/ 01/ 1994	20	SMAN 1 Darma	IPA			KARYA WAN	Spanduk	Jalan cileungsir - sukadana Kec. Darma Kab. Kuningan	1.000.000 s/d 1.900.000
4	14152 21003	Putri Rahmawati	P	Ilmu Gizi	Kuningan	01/ 01/ 1994	20	SMAN 1 Kadugede	IPA	Negeri	43,72	REGULER	Brosur	Kadugede Kuningan	1.000.000 s/d 1.900.000
5	14152 21004	Ratna Sadi'ah	P	Ilmu Gizi								KARYA WAN		Desa Sembawa RT/RW Kec Jalaksana Kab. Kuningan	1.000.000 s/d 1.900.000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
324	14153 41001	IMAM HILMY NAUFAL	L	TP	Kuningan									Cigugur	

Data lengkap mahasiswa baru dari tahun 2014 terlampir (Lampiran 2).

Tabel 4.3 merupakan jumlah data kosong pada setiap atribut pada tahun 2014 sampai dengan 2019.

Tabel 4.3 Data Kosong Pada Setiap Atribut Tahun 2014-2019

Atribut	2014	2015	2016	2017	2018	2019
NPM	0	0	0	0	0	0
Nama Mahasiswa	0	0	0	0	0	0
L/P	0	0	0	0	0	0
Prodi	0	0	0	0	0	0
Tempat Lahir	21	9	1	0	4	3
Tanggal Lahir	5	5	9	4	5	4
Umur	6	5	9	4	3	4
Asal Sekolah	10	4	8	4	0	4
Program	33	12	19	16	17	8
Status Sekolah	53	34	15	10	8	5
Nilai UN	62	42	10	20	12	
KELAS	0	0	0	0	0	0
Media	17	20	16	15	16	12
Alamat	2	4	7	4	0	3
Gaji Orang Tua	8	5	9	10	7	2

4.2 *Selection* Data

Selection data digunakan untuk memilih data apa saja yang berpengaruh dan tidak berpengaruh dengan strategi promosi. Tabel 4.4 merupakan atribut-atribut pada data mahasiswa baru yang akan digunakan dan Tabel 4.5 merupakan atribut-aribut yang akan dihilangkan. Informasi didapat berdasarkan wawancara dengan bagian PUSINFO terlampir (Lampiran 1).

Tabel 4.4 Pemilihan Atribut yang Digunakan

Atribut Yang Digunakan		Keterangan
1.	Umur	Untuk mengetahui sekolah yang langsung melanjutkan kuliah
2.	Asal Sekolah	Untuk mengetahui sasaran promosi yang tepat
3.	Media	Untuk media promosi yang digunakan

Atribut Yang Digunakan		Keterangan
4.	Kecamatan	Untuk penempatan spanduk yang tepat
5.	Gaji Orang Tua	Untuk mengetahui gaji orang tua menengah ke bawah

Tabel 4.5 Pemilihan Atribut yang Dihilangkan

Atribut Yang Dihilangkan		Keterangan
1.	No	Memiliki nilai unik
2.	NPM	Memiliki nilai unik
3.	Nama	Memiliki nilai unik
4.	L/P	Laki-laki maupun perempuan bisa masuk ke UNISA
5.	Prodi	Semua prodi dilakukan sosialisasi secara bersamaan
6.	Tempat Lahir	Dianggap sama dengan alamat
7.	Tanggal Lahir	Dianggap sama dengan Umur
8.	Program	Semua Program Studi dibuka untuk semua program
9.	Status Sekolah	Dianggap sama dengan asal sekolah
10.	Nilai UN	Berapapun Nilai UN dapat masuk ke UNISA
11	Kelas	Tidak ada perbedaan khusus antara kelas karyawan dan Reguler, yang membedakan hanya biaya dan waktu perkuliahan

4.3 *Preprocessing* Data

Salah satu tahapan *preprocessing* data adalah pembersihan data. Tahapan ini dilakukan dengan cara memperbaiki atau menghilangkan data yang kosong, tidak konsisten, memperbaiki kesalahan pada data seperti kesalahan dalam pengetikan tidak lengkap, dan menghilangkan duplikasi data. Pengisian nilai kosong dapat menggunakan modus,*mean*, random. Modus merupakan pengisian data diisi dengan data yang paling banyak muncul seperti pada Tabel 4.6.

Tabel 4.6 Pengisian Data Kosong dengan Modus

No.	Asal Sekolah	Media
59.	SMKN 2 Kuningan	Rekomendasi
74.	SMKN 2 Kuningan	Rekomendasi
137.	SMKN 2 Kuningan	Brosur
140.	SMKN 2 Kuningan	Brosur
184.	SMKN 2 Kuningan	Spanduk
191.	SMKN 2 Kuningan	Brosur
202.	SMKN 2 Kuningan	Brosur
224.	SMKN 2 Kuningan	*Brosur

No.	Asal Sekolah	Media
231.	SMKN 2 Kuningan	Rekomendasi

Tabel 4.6 pada *record* 224 mempunyai satu data kosong pada atribut “Media”, kemudian akan digunakan penggunaan modus yang diurutkan berdasarkan Asal Sekolah abjad A-Z dari semua data pada atribut “Media” tahun 2014 sehingga data kosong diisi dengan media “Brosur”. Selain “Media”, atribut “Gaji Orang Tua” dan “Alamat” juga menggunakan modus. *Mean* merupakan pengisian data kosong yang diisi dengan rata-rata seperti pada Tabel 4.7.

Tabel 4.7 Pengisian Data Kosong dengan *Mean*

No.	Umur	Asal Sekolah
8.	22	SMAN 1 Ciawigebang
21.	21	SMAN 1 Ciawigebang
34.	19	SMAN 1 Ciawigebang
38.	*21	SMAN 1 Ciawigebang
44.	19	SMAN 1 Ciawigebang
87.	22	SMAN 1 Ciawigebang
143.	19	SMAN 1 Ciawigebang
155.	19	SMAN 1 Ciawigebang
163.	19	SMAN 1 Ciawigebang
212.	22	SMAN 1 Ciawigebang
215.	24	SMAN 1 Ciawigebang
248.	21	SMAN 1 Ciawigebang
252.	25	SMAN 1 Ciawigebang
259.	19	SMAN 1 Ciawigebang
271.	24	SMAN 1 Ciawigebang
322.	24	SMAN 1 Ciawigebang

Tabel 4.7 pada *record* 38 mempunyai data kosong pada atribut “Umur”, kemudian akan digunakan penggunaan rata-rata yang diurutkan berdasarkan Asal Sekolah “SMAN 1 Ciawigebang” tahun 2014 sehingga nilai rata-ratanya 20,7 yang dibulatkan menjadi 21. *Random* merupakan pengisian data kosong yang diisi secara random seperti pada Tabel 4.8.

Tabel 4.8 Pengisian Data Kosong dengan *Random*

No.	Asal Sekolah	Media
43.	MAN Kota Tegal	*Internet

Tabel 4.8 pada *record* 43 mempunyai data kosong pada aribut “Media”, kemudian akan digunakan penggunaan random karena mahasiswa yang asal sekolahnya MAN Kota Tegal hanya ada satu, sehingga data kosong diisi dengan data media “Internet”. Contoh pengisian data tidak konsisten seperti pada Tabel 4.9.

Tabel 4.9 Perbaikan Data Tidak Konsisten

No.	Asal Sekolah
29.	SMA Kosgoro Kuningan
129.	*SMA Kosgoro Kuningan
173.	SMA Kosgoro Kuningan
266.	SMA Kosgoro Kuningan

Tabel 4.8 yang telah diurutkan berdasarkan asal sekolah sesuai abjad A-Z, kolom asal sekolah pada *record* no 129 diisi dengan “SMA Kosgoro”, sedangkan pada data asal sekolah lain diisi dengan “SMA Kosgoro Kuningan” sehingga asal sekolah pada *record* no 129 harus diubah agar datanya menjadi konsisten. Sehingga data akan diperbaiki menjadi “SMA Kosgoro Kuningan”. Contoh perbaikan data dalam salah pengetikan data seperti Tabel 4.10.

Tabel 4.10 Perbaikan Data dalam Pengetikan Data

No.	Asal Sekolah
215.	SMAN 1 Ciawigebang
248.	SMAN 1 Ciawigebang
252.	SMAN 1 Ciawigebang
259.	*SMAN 1 Ciawigebang
271.	SMAN 1 Ciawigebang

Gambar 4.10 telah diurutkan bedasarkan asal sekolah sesuai abjad A-Z, kolom asal sekolah pada *record* no 259 kolom asal sekolah diisi dengan “SMAN 1 Ciawigbang”, sedangkan pada data asal sekolah lain diisi dengan “SMAN 1 Ciawigbang” sehingga asal sekolah pada *record* no 259 harus diubah agar datanya menjadi konsisten. Sehingga data diperbaiki menjadi “SMAN 1 Ciawigebang”. Apabila data kosong dalam satu *record* masih banyak, maka dapat dilakukan penghapusan satu *record* seperti pada Tabel 4.11. yang mempunyai lima data kosong.

Tabel 4.11 Menghilangkan Data Yang Mempunyai Banyak Data Kosong

No	Umur	Asal Sekolah	Media	Alamat	Gaji Ortu
57.					

Setelah tidak ada data yang kosong, maka data mahasiswa baru diintegrasikan dengan cara menggabungkan tabel mahasiswa dari tahun 2014 sampai dengan tahun 2019 menjadi satu tabel. Jumlah total hasil akhir *record* data mahasiswa baru setelah dilakukan proses pembersihan data adalah 1939 *record*. Hasil integrasi tabel seperti pada Tabel 4.12.

Tabel 4.12 Hasil Integrasi Tabel Tahun 2014 sampai 2019

No	Umur	Asal Sekolah	Media	Alamat	Gaji Orang Tua (Rp)
1.	19	SMAN 1 Cigugur	Brosur	Kuningan Kab.Kuningan	2.000.000 s/d 2.900.000
2.	19	SMK YAMSIK Kuningan	Brosur	Kuningan Kab.Kuningan	3.000.000 s/d 3.900.000
3.	20	SMAN 1 Darma	Spanduk	Jalan cileungsir - sukadana Kec. Darma Kab. Kuningan	1.000.000 s/d 1.900.000
4.	20	SMAN 1 Kadugede	Brosur	Kadugede Kuningan	1.000.000 s/d 1.900.000
5.	20	MAN 1 Kuningan	Brosur	Desa Sembawa RT/RW Kec Jalaksana Kab. Kuningan	1.000.000 s/d 1.900.000
...	...	...	...	...	...
1939.	20	SMAN 1 Kuningan	Brosur	Kuningan	3.000.000 s/d 3.900.000

Data mahasiswa hasil integrasi terlampir (Lampiran 3).

4.4 Transformasi Data

Atribut asal sekolah dan alamat akan ditransformasikan berdasarkan jarak dari asal sekolah atau alamat mahasiswa ke UNISA Kuningan. Untuk alamat akan di ambil data Kecamatan sesuai kebutuhan Universitas. Tabel 4.13 menampilkan hasil atribut alamat yang sudah dipisah berdasarkan jarak dari kecamatan ke UNISA Kuningan.

Tabel 4.13 Transformasi Alamat menjadi Kecamatan

No.	Alamat	Hasil Transformasi
1.	Kuningan Kab.Kuningan	5.8
2.	Kuningan Kab.Kuningan	5.8
3.	Jalan cileungsir - sukadana Kec. Darma Kab. Kuningan	14.1
4.	Kadugede Kuningan	5.9

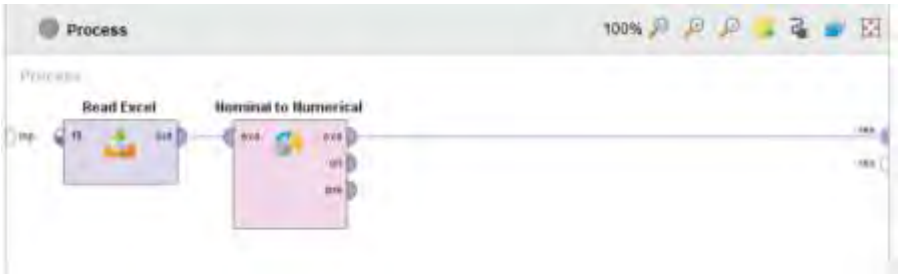
No.	Alamat	Hasil Transformasi
5.	Desa Sembawa RT/RW Kec Jalaksana Kab. Kuningan	7.5
...	...	...
1939.	Kel. Winduhaji Kec. Kuningan Kab. Kuningan Kec. Kuningan Kab.Kuningan	5.8

Tabel 4.14 merupakan Atribut “Gaji Orang Tua” yang ditransformasikan menggunakan rata-rata.

Tabel 4.14 Transformasi Gaji Orang Tua Menggunakan *Means*

No	Gaji Orang Tua (Rp)	Hasil Transformasi (Rp)
1.	1.000.000 s/d 1.900.000	1.450.000
2.	2.000.000 s/d 2.900.000	2.450.000
3.	3.000.000 s/d 3.900.000	3.450.000
4.	4.000.000 s/d 4.900.000	4.450.000
5.	5.000.000 s/d 5.900.000	5.450.000
6.	6.000.000 s/d 6.900.000	6.450.000

Tahapan selanjutnya adalah dengan mentransformasikan data dari data yang berbentuk nominal menjadi numerik pada atribut Media. Hal ini dilakukan karena Algoritma yang digunakan dan validasi dunn *Index* hanya dapat dilakukan dengan data yang berbentuk numerik. Transformasi dilakukan dengan menggunakan aplikasi *rapidminer* yaitu dengan menggunakan operator *Read Excel* karena data berupa *excel* , dan *nominal to numerical* untuk merubah data yang berbentuk teks menjad numerik seperti pada Gambar 4.1.



Gambar 4.1
Penambahan Operator *Read Excel* dan Nominal to *Numerical*

Berdasarkan Gambar 4.1 setelah dilakukan proses *running* data akan otomatis berubah menjadi angka dengan inisialisasi angka 0 sampai tidak terhingga. Inisialisasi

dilakukan sesuai dengan urutan muculnya data sehingga Media dapat di transformasikan seperti Tabel 4.15.

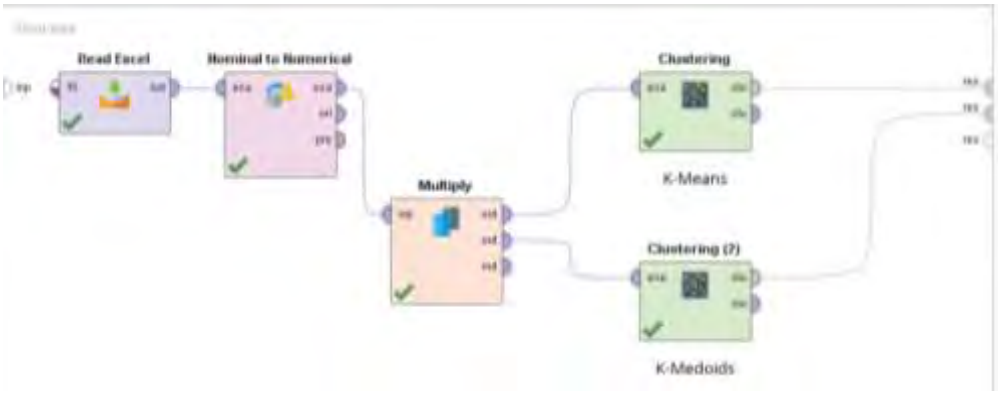
Tabel 4.15 Transformasi Data Media

No.	Media	Hasil Transformasi
1.	Brosur	0
2.	Spanduk	1
3.	Rekomendasi	2
4.	Internet	3

Sumber : Rapidminer

4.5 *Data Mining*

Untuk menggali informasi data mahasiswa baru yang diambil dari Tabel 4.12 Tahun 2014 sampai dengan Tahun 2019 akan menggunakan *tools Rapidminner*. Pada tahapan ini diperlukan operator algoritma yang akan digunakan yaitu algoritma *K-Means* dan Algoritma *K-Medoids* seperti Gambar 4.2.



Gambar 4.2
Penambahan Operator Multiply dan Algoritma

Pada Gambar 4.2 terdapat operator *multiply* yang berfungsi untuk menghubungkan Algoritma *K-Means* pada *cluste*ring pertama dan *K-Medoids* pada *clustering* kedua. Parameter *clustering K-Means* yang digunakan dalam penelitian ini seperti pada Tabel 4.16.

Tabel 4.16 Parameter *clustering K-Means*

Parameter	Value
K	2 sampai dengan 10
Max runs	10 (default)
Measure type	Numerical measurement

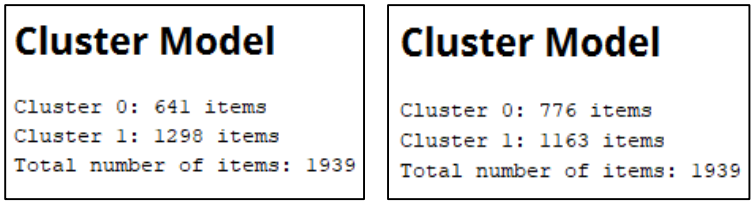
Numerical measure	EuclideanDistance
Max optimization step	100 (default)

Parameter *clustering K-Medoids* yang digunakan dalam penelitian ini seperti pada Tabel 4.17.

Tabel 4.17 Parameter *Clustering K-Medoids*

Parameter	Value
K	2 sampai dengan 10
Max runs	10 (default)
Max Optimization Step	100
Measure type	Numerical Measurement
Numerical measure	EuclideanDistance
Max optimization step	100 (default)

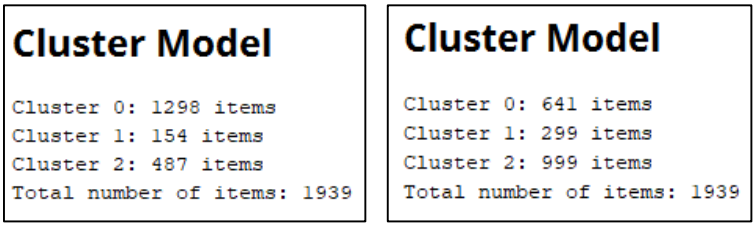
Perbandingan *Cluster* Model dari K=2 sampai dengan K=10 seperti pada Gambar 4.3 sampai dengan 4.12.



(a) (b)

Gambar 4.3
Cluster model K = 2 pada *K-Means* (a) dan *K-Medoids* (b)

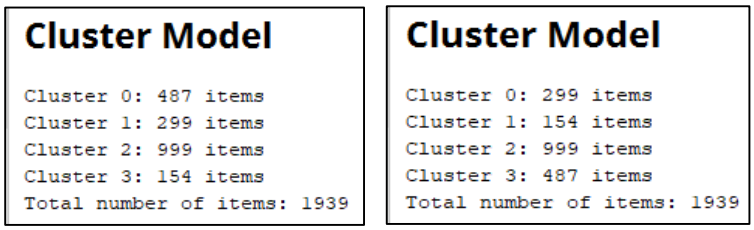
Berdasarkan Gambar 4.3, jumlah data pengelompokkan data K=2 pada Algoritma *K-Means* di *cluster* 0 berjumlah 641 *record* dan *cluster* 1 berjumlah 1298 *record*. Sedangkan algoritma *K-Medoids* K=2 pada *cluster* 0 sebesar 776 *record* dan *cluster* 1 berjumlah 1163 *record*.



(a) (b)

Gambar 4.4
Cluster model K = 3 pada *K-Means* (a) dan *K-Medoids* (b)

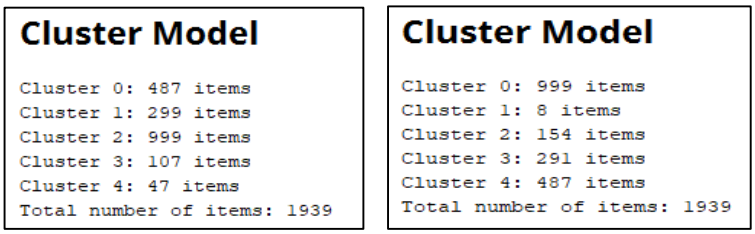
Berdasarkan Gambar 4.4, jumlah data pengelompokkan data K=3 pada Algoritma *K-Means* di *cluster* 0 berjumlah 1298 *record*, *cluster* 1 berjumlah 154 *record* dan *cluster* 3 berjumlah 487 *record*. Sedangkan algoritma *K-Medoids* K=3 pada *cluster* 0 sebesar 641 *record* , *cluster* 1 berjumlah 299 *record* dan *cluster* 2 berjumlah 1939 *record*.



(a) (b)

Gambar 4.5
Cluster model K = 4 pada *K-Means* (a) dan *K-Medoids* (b)

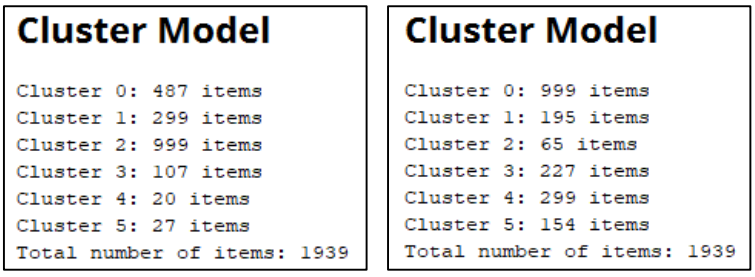
Berdasarkan Gambar 4.5, jumlah data pengelompokkan data K=4 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 999 *record* dan *cluster* 3 berjumlah 154 *record*. Sedangkan algoritma *K-Medoids* K=4 pada *cluster* 0 sebesar 299 *record* , *cluster* 1 berjumlah 154 *record*, *cluster* 2 berjumlah 999 *record* dan *cluster* 3 berjumlah 487 *record*.



(a) (b)

Gambar 4.6
Cluster model K = 5 pada *K-Means* (a) dan *K-Medoids* (b)

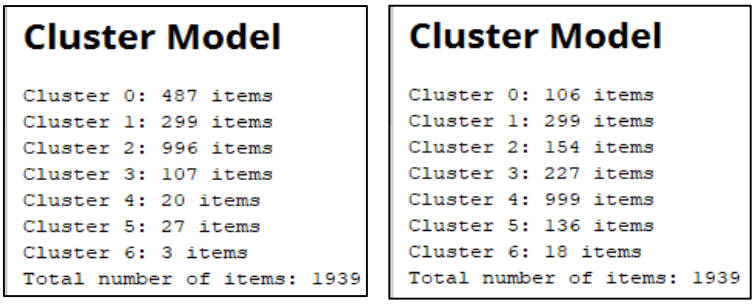
Berdasarkan Gambar 4.6, jumlah data pengelompokkan data K=5 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 999 *record*, *cluster* 3 berjumlah 107 *record* dan *cluster* 4 berjumlah 47 *record*. Sedangkan algoritma *K-Medoids* K=5 pada *cluster* 0 sebesar 999 *record* , *cluster* 1 berjumlah 8 *record*, *cluster* 2 berjumlah 154 *record, cluster* 3 berjumlah 291 *record* dan *cluster* 4 berjumlah 487 *record*.



(a) (b)

Gambar 4.7
Cluster model K = 6 pada *K-Means* (a) dan *K-Medoids* (b)

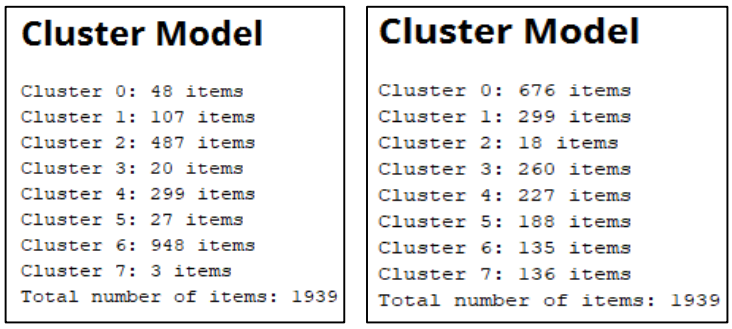
Berdasarkan Gambar 4.7, jumlah data pengelompokkan data K=6 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 999 *record*, *cluster* 3 berjumlah 107 *record, cluster* 4 berjumlah 20 *record,* dan *cluster* 5 berjumlah 27 *record*. Sedangkan algoritma *K-Medoids* K=6 pada *cluster* 0 sebesar 999 *record* , *cluster* 1 berjumlah 195 *record*, *cluster* 2 berjumlah 65 *record, cluster* 3 berjumlah 227 *record, cluster* 4 berjumlah 299 *record* dan *cluster* 5 berjumlah 154 *record*.



(a) (b)

Gambar 4.8
Cluster model K = 7 pada *K-Means* (a) dan *K-Medoids* (b)

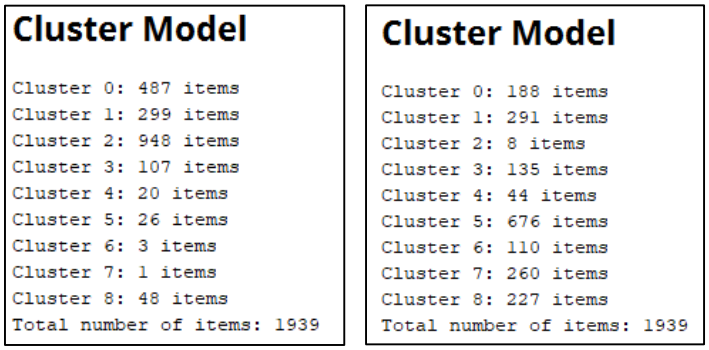
Berdasarkan Gambar 4.8, jumlah data pengelompokkan data K=7 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 996 *record*, *cluster* 3 berjumlah 107 *record, cluster* 4 berjumlah 20 *record, cluster* 5 berjumlah 27 *record* dan *cluster* 6 berjumlah 3 *record*. Sedangkan algoritma *K-Medoids* K=7 pada *cluster* 0 sebesar 106 *record* , *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 154 *record, cluster* 3 berjumlah 227 *record, cluster* 4 berjumlah 999 *record, cluster* 5 berjumlah 136 *record* dan *cluster* 6 berjumlah 18 *record*.



(a) (b)

Gambar 4.9
Cluster model K = 8 pada *K-Means* (a) dan *K-Medoids* (b)

Berdasarkan Gambar 4.9, jumlah data pengelompokkan data K=8 pada Algoritma *K-Means* di *cluster* 0 berjumlah 48 *record*, *cluster* 1 berjumlah 107 *record*, *cluster* 2 berjumlah 487 *record*, *cluster* 3 berjumlah 20 *record, cluster* 4 berjumlah 299 *record, cluster* 5 berjumlah 27 *record, cluster* 6 berjumlah 948 *record,* dan *cluster* 7 berjumlah 3 *record*. Sedangkan algoritma *K-Medoids* K=8 pada *cluster* 0 berjumlah 676 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 18 *record, cluster* 3 berjumlah 260 *record, cluster* 4 berjumlah 227 *record, cluster* 5 berjumlah 188 *record, cluster* 6 berjumlah 135 *record* dan *cluster* 7 berjumlah 136 *record*.

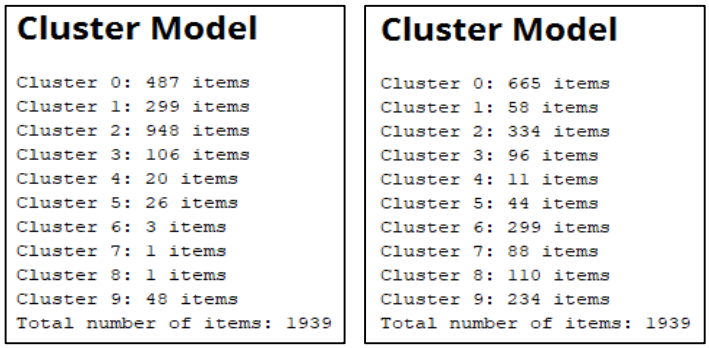


(a) (b)

Gambar 4.10
Cluster model K = 9 pada *K-Means* (a) dan *K-Medoids* (b)

Berdasarkan Gambar 4.10, jumlah data pengelompokkan data K=9 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 948 *record*, *cluster* 3 berjumlah 107 *record, cluster* 4 berjumlah 20 *record, cluster* 5 berjumlah 26 *record, cluster* 6 berjumlah 3 *record, cluster* 7 berjumlah 1 *record,* dan *cluster* 8 berjumlah 48 *record*. Sedangkan algoritma *K-Medoids* K=9 pada *cluster* 0 sebesar 188 *record* , *cluster* 1 berjumlah 291 *record*, *cluster* 2 berjumlah 8 *record,*

cluster 3 berjumlah 135 *record, cluster* 4 berjumlah 44 *record, cluster* 5 berjumlah 676 *record, cluster* 6 berjumlah 110 *record*, *cluster* 7 berjumlah 260 *record* dan *cluster* 8 berjumlah 227 *record*.



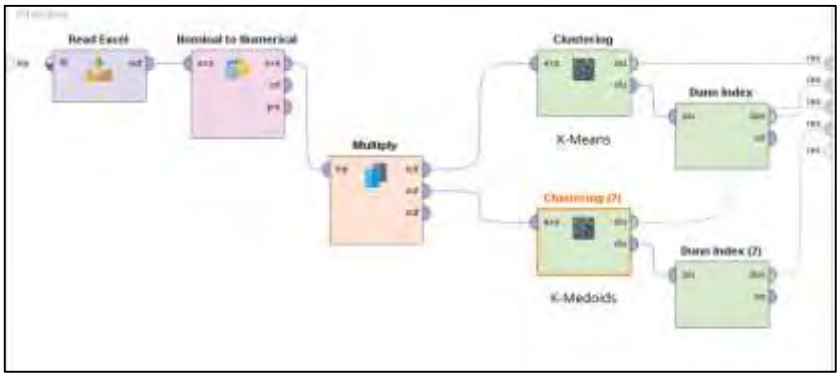
(a) (b)

Gambar 4. 11
Cluster model K = 10 pada *K-Means* (a) dan *K-Medoids* (b)

Berdasarkan Gambar 4.11, jumlah data pengelompokkan data K=10 pada Algoritma *K-Means* di *cluster* 0 berjumlah 487 *record*, *cluster* 1 berjumlah 299 *record*, *cluster* 2 berjumlah 948 *record*, *cluster* 3 berjumlah 106 *record, cluster* 4 berjumlah 20 *record, cluster* 5 berjumlah 26 *record, cluster* 6 berjumlah 3 *record, cluster* 7 berjumlah 1 *record, cluster* 8 berjumlah 1 *record*, dan *cluster* 9 berjumlah 48 *record.* Sedangkan algoritma *K-Medoids* K=9 pada *cluster* 0 sebesar 665 *record* , *cluster* 1 berjumlah 58 *record*, *cluster* 2 berjumlah 334 *record, cluster* 3 berjumlah 96 *record, cluster* 4 berjumlah 11 *record, cluster* 5 berjumlah 44 *record, cluster* 6 berjumlah 299 *record*, *cluster* 7 berjumlah 88, *cluster* 8 berjumlah 110 *record* dan *cluster* 9 berjumlah 234 *record*.

4.6 *Pattern Evaluation*

Setelah mendapatkan pola-pola unik pada percobaan K=2 sampai dengan K=10, maka untuk mengetahui mana K terbaik adalah dengan menggunakan operator validasi dunn *Index* seperti Gambar 4.12



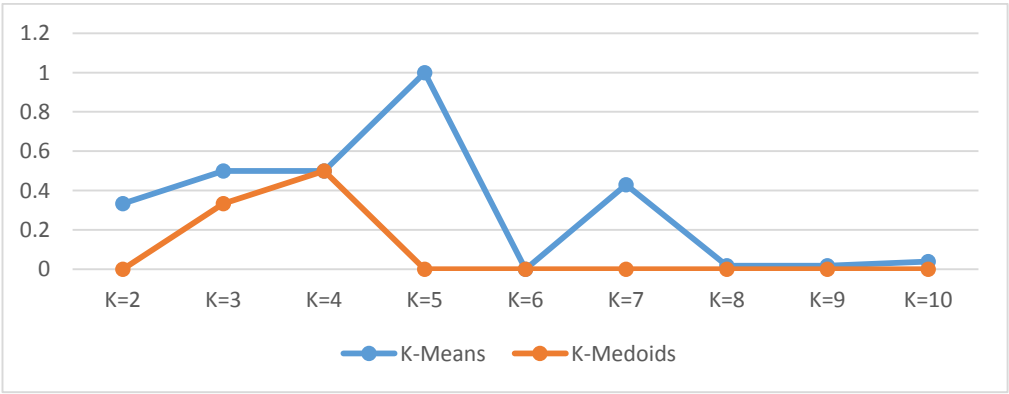
Gambar 4.12
Penambahan operator validasi dunn *Index*

Setelah melakukan percobaan terhadap parameter tersebut pada Tabel 4.18 diperoleh Nilai Dunn *Index* dan lama proses *running.*

Tabel 4.18 *Cluster K-Means* dan *K-Medoids* terhadap Nilai Dunn *Index*

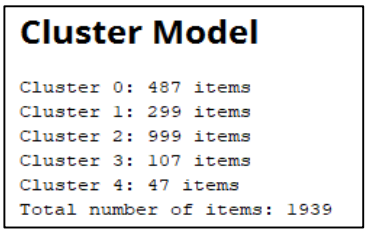
Cluster	***K-Means***	**Waktu**	***K-Medoids***	**Waktu**
2	0.333	1s	∞	38s
3	0.499	1s	0.333	42s
4	0.499	1s	0.499	49s
5	0.999	1s	∞	54s
6	∞	1s	∞	1m 01s
7	0,430	1s	∞	1m 04s
8	0.017	1s	∞	1m 10s
9	0.017	1s	∞	1m 17s
10	0.039	1s	∞	1m 24s

Gambar 4.13 merupakan grafik berdasarkan Tabel 4.19 sehingga nilai dunn *Index* mudah untuk dipahami.



Gambar 4.13
Grafik *Cluster K-Means* dan *K-Medoids* terhadap Nilai Dunn *Index*

Berdasarkan (Chowdary, 2014) yang menyatakan bahwa semakin besar indeksnya, semakin baik hasil pengelompokannya, maka dilihat dari Gambar 4.13, nilai K yang optimum pada algoritma *K-Means* adalah *cluster* ke 5 dan pada algoritma *K-Medoids* *cluster* ke 4. Kemudian dilihat kembali nilai dunn *Index* yang paling tinggi antara k=5 pada algoritma *K-Means* dan k=4 pada algoritma *K-Medoids*. Sehingga dapat disimpulkan bahwa algoritma *K-Means* lebih baik dengan nilai dunn *Index* 0,999. Adapun dalam pengolahannya seperti pada Tabel 4.4, *K-Means* juga lebih baik karena hanya membutuhkan waktu rata-rata 1 detik sedangkan pengolahan data pada *K-Medoids* membutuhkan waktu rata-rata 1 menit 05 detik yang artinya makin tinggi pengelompokkan yang ditentukan, maka pengolahan data akan semakin lama. Sehingga pada penelitian ini algoritma *K-Means* lebih baik daripada algoritma *K-Medoids* dengan K=5. *Cluster Model* dan *Centroid Table K-Means* dengan k=5 seperti Gambar 4.14.



Gambar 4.14
Cluster Model Algoritma *K-Means* pada K = 5

Gambar 4.14 menjelaskan jumlah data mahasiswa baru pada setiap *cluster*nya yang terbagi menjadi 5 *cluster*, dimana *cluster* 0 dengan jumlah 487 data, *cluster* 1 dengan 299 data, *cluster* 2 dengan 999 data, *cluster* 3 dengan 107 data dan *cluster* 4 dengan 47 data. Tabel 4.19 merupakan hasil perhitungan antara jarak *cluster* dengan centroid menggunakan *Rapidminer*.

Tabel 4.19 Hasil Perhitungan Jarak *Cluster* dengan *Centroid*

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Media = Internet	0,035	0,097	0,280	0,196	0,191
Media = Rekomendasi	0,392	0,247	0,116	0,234	0,426
Media = Brosur	0,419	0,515	0,302	0,121	0,064
Media = Spanduk	0,154	0,140	0,301	0,449	0,319
Umur	19,786	22,903	23,100	19,290	19,340
Jarak Sekolah	25,217	19,837	35,233	66,870	101,487

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gaji Orang Tua	3.000.000 s/d 3.900.000	1.450.000	2.450.000	4.450.000	5.875.531,9
Jarak Kecamatan	24,821	16,009	32,809	64,469	99,489

Sumber : Rapidminer

4.7 *Interpretation Knowledge*

Penjabaran dari *Cluster* Model pada Gambar 4.14 .dengan nilai K=5 pada Algoritma *K-Means* seperti Tabel 4.20. yang berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data mahasiswa pada setiap atribut.seperti pada Tabel 4.20 sampai Tabel 4.24.

Tabel 4.20 Hasil Analisis *Cluster* 0

Cluster 0									
Jarak Sekolah/Km									
17,8	29		27,7	4		65,6	2	18,8	1
0,1	23		16,9	4		18,9	2	113,0	1
2,5	21		12,5	4		11,4	2	13,9	1
6,3	20		12,7	4		45,0	2	315,0	1
3,2	19		17,4	4		38,2	1	46,8	1
4,5	17		16,8	3		44,2	1	11,6	1
5,5	14		17,5	3		33,1	1	36,4	1
4,6	12		30,3	3		37,5	1	28,6	1
15,3	11		19,7	3		42,4	1	4,8	1
18,6	11		28,4	3		235,0	1	32,8	1
5,1	10		3,9	3		47,6	1	35,7	1
11,0	9		28,8	3		15,7	1	13,1	1
38,3	9		231,0	3		47,8	1	112,0	1
9,6	9		48,4	3		28,3	1	35,9	1
9,5	8		33,2	3		56,4	1	15,1	1
14,2	8		14,1	3		16,5	1	20,6	1
5,0	7		14,5	3		62,1	1	14,0	1
25,0	7		11,1	3		16,4	1	38,1	1
21,2	7		13,4	2		33,4	1	136,0	1
15,2	7		282,0	2		251,0	1	19,2	1
4,2	6		11,2	2		70,2	1	62,8	1
16,0	6		213,0	2		699,0	1	9,0	1
2,6	6		132,0	2		42,8	1	36,2	1
2,1	6		6,2	2		6,0	1	17,3	1
35,8	6		35,3	2		39,6	1	239,0	1
2,0	6		13,8	2		18,4	1	26,2	1
2,7	6		22,1	2		378,0	1	25,7	1
22,2	6		23,5	2		6,9	1	2,2	1
5,2	5		47,7	2		48,5	1	16,2	1
48,0	5		188,0	2		55,1	1	14,9	1

Cluster 0										
66,9	5		8,3	2		1,3	1		8,6	1
16,3	5		17,7	2		8,9	1			
10,1	4		15,0	2		375,0	1			
Jarak Kecamatan/Km										
5,8	193		35,8	6		37,0	3		33,3	1
18,4	30		30,1	6		43,4	2		251,0	1
16,4	19		7,4	6		61,2	2		699,0	1
44,4	17		1,7	6		225,0	2		236,0	1
14,1	14		22,7	5		68,1	2		378,0	1
5,9	14		24,8	5		62,8	2		55,2	1
9,0	14		20,8	5		301,0	2		367,0	1
7,5	12		50,1	4		33,4	2		101,0	1
11,1	11		37,5	4		35,4	2		315,0	1
46,9	9		22,0	4		67,7	2		192,0	1
19,2	8		22,9	3		18,3	1		119,0	1
22,4	8		17,1	3		234,0	1		148,0	1
17,6	8		14,0	3		29,0	1		53,0	1
10,4	7		37,6	3		62,9	1		47,0	1
23,2	7		26,2	3		276,0	1			
27,5	6		17,0	3		280,0	1			
Media										
Brosur	204		Rekomendasi	191		Spanduk	75		Internet	17
Umur						**Rata-Rata Gaji Orang Tua**				
19						3.000.000 s/d 3.900.000				

Berdasarkan Tabel 4.20 *Cluster* 0 memiliki karakteristik calon mahasiswa berumur 19 tahun dengan rata-rata gaji orang tua sebesar Rp. 3.000.000 s/d 3.900.000. Jarak sekolah ±18,0 Km paling banyak memberikan lulusan dengan jumlah 29 calon mahasiswa dengan persentase 6%. Kecamatan yang mempunyai jarak ±5,8 Km mempunyai persentase paling besar dengan 39,6%. Brosur dan Rekomendasi sangat berpegaruh dengan perbedaan persentase 3,3%.

Tabel 4.21 Hasil Analisis *Cluster* 1

Cluster 1										
Jarak Sekolah/Km										
3,2	61		27,7	3		16,3	2		29,4	1
0,1	26		19,7	3		47,2	2		22,2	1
6,3	15		11,1	3		14,1	2		7,5	1
42,4	12		1,7	3		39,6	1		16,4	1
15,2	11		5,0	3		6,9	1		284,0	1
18,0	11		2,2	3		146,0	1		81,4	1
2,5	9		15,3	2		38,2	1		66,9	1
4,6	8		18,9	2		93,1	1		37,4	1
9,6	8		5,1	2		39,5	1		62,1	1

Cluster 1									
18,6	7		40,0	2		48,0	1	235,0	1
35,8	7		44,2	2		36,2	1	12,5	1
11,0	6		16,0	2		48,4	1	18,4	1
21,2	6		5,5	2		15,5	1	39,7	1
4,5	6		38,3	2		6,0	1	1,9	1
4,2	5		14,2	2		79,3	1	23,5	1
33,2	5		2,6	2		56,4	1	45,3	1
9,5	5		109,0	2		103,0	1	129,0	1
16,8	4		9,0	2		250,0	1		
2,0	4		282,0	2		43,3	1		
Jarak Kecamatan/Km									
1,7	159		37,6	3		27,5	1	276,0	1
14,1	27		24,8	2		172,0	1	22,7	1
9,0	13		44,4	2		16,4	1	7,6	1
18,4	10		158,0	2		10,4	1	76,7	1
5,9	9		7,4	2		40,8	1	14,0	1
23,2	9		37,5	2		18,3	1	61,4	1
43,4	9		47,2	2		280,0	1	61,2	1
35,8	7		22,4	2		56,4	1	234,0	1
7,5	3		37,0	2		104,0	1	17,6	1
5,8	3		6,0	1		30,1	1	38,9	1
11,1	3		42,8	1		249,0	1		
17,1	3		47,0	1		26,1	1		
Media									
Brosur	154		Rekomendasi	74		Spanduk	42	Internet	29
Rata-Rata Umur						**Rata-Rata Gaji Orang Tua**			
23						1.450.000			

Berdasarkan Tabel 4.21 *Cluster* 1 memiliki karakteristik calon mahasiswa berumur 23 tahun dengan rata-rata gaji orang tua sebesar Rp. 1.450.000. Jarak sekolah ±3,2 Km paling banyak memberikan lulusan dengan jumlah 61 calon mahasiswa dengan persentase 20,4%. Kecamatan yang mempunyai jarak ±1,7 Km mempunyai persentase paling besar dengan 53,2%. Brosur sangat berpegaruh dengan persentase 51,5%

Tabel 4.22 Hasil Analisis *Cluster* 2

Cluster 2									
Jarak Asal Sekolah/Km									
6,3	52		109,0	4		586,0	2	36,3	1
42,4	43		17,4	4		90,1	2	125,0	1
11,0	39		6,9	4		46,2	2	13,1	1
14,2	39		27,7	4		6,2	2	35,3	1
4,6	38		5,9	4		56,9	2	197,0	1
9,6	37		28,8	4		171,0	2	74,7	1
18,6	29		365,0	4		159,0	2	11,3	1
2,5	29		16,5	4		11,2	2	5,8	1

Cluster 2									
5,5	29		14,1	4		272,0	2		19,0	1
0,1	28		80,2	4		10,1	2		16,2	1
18,0	26		17,2	4		113,0	1		46,8	1
21,2	24		45,0	4		37,5	1		103,0	1
3,2	21		6,0	3		236,9	1		62,8	1
15,2	21		19,7	3		223,0	1		186,0	1
4,5	20		39,6	3		45,1	1		108,0	1
9,5	19		237,0	3		360,0	1		146,0	1
35,8	17		44,2	3		33,1	1		391,0	1
11,1	17		25,0	3		34,9	1		36,6	1
4,2	16		235,0	3		174,0	1		271,0	1
38,3	16		250,0	3		18,0	1		24,2	1
2,6	16		28,3	3		1046,0	1		6,1	1
16,0	15		17,1	3		47,6	1		42,3	1
16,9	14		11,6	3		37,4	1		46,9	1
22,2	12		9,0	3		220,0	1		1861,0	1
16,3	11		39,5	3		330,0	1		262,0	1
5,1	11		15,7	3		350,0	1		31,7	1
2,0	11		55,6	3		220,1	1		23,8	1
12,7	10		132,0	3		17,9	1		2,2	1
16,8	10		16,6	3		7,5	1		4,3	1
14,5	9		18,7	2		66,7	1		272,1	1
33,2	9		40,0	2		30,3	1		33,8	1
38,1	8		54,5	2		246,0	1		19,1	1
12,5	8		36,2	2		65,6	1		45,2	1
15,3	7		378,0	2		20,6	1		81,4	1
48,4	7		48,0	2		1556,0	1		26,2	1
17,7	6		16,4	2		80,6	1		55,1	1
2,7	6		21,0	2		175,0	1		284,0	1
2,1	6		29,4	2		62,1	1		3,5	1
249,0	6		23,5	2		54,3	1		4,8	1
18,9	5		38,2	2		66,9	1		44,3	1
15,5	5		15,4	2		28,4	1		378,1	1
15,0	5		17,0	2		32,6	1		54,4	1
13,4	5		8,3	2		93,1	1			
5,0	5		49,7	2		41,6	1			
5,2	4		242,0	2		13,9	1			
Jarak Kecamatan/Km										
5,9	140		22,9	8		234,0	2		86,1	1
16,4	88		27,5	7		377,0	2		193,0	1
5,8	77		22,4	7		237,0	2		365,0	1
18,4	70		47,0	7		47,4	2		42,9	1
9,0	70		37,6	7		25,4	2		36,2	1
14,1	68		42,8	6		47,6	2		37,2	1
11,1	42		30,1	6		7,4	2		36,1	1
43,4	31		1,7	6		92,5	2		127,0	1
24,8	29		17,1	6		137,0	2		50,4	1
19,2	22		62,8	6		50,2	2		175,0	1
10,4	22		158,0	5		46,5	1		117,0	1

Cluster 2									
23,2	21		26,2	4		101,0	1		147,0	1
7,5	20		89,8	4		40,8	1		168,0	1
35,8	19		61,2	4		68,1	1		47,3	1
44,4	15		46,9	3		35,4	1		261,0	1
17,6	13		37,5	3		18,3	1		1601,0	1
22,7	13		29,0	3		172,0	1		232,0	1
20,8	12		243,0	3		1046,0	1		34,5	1
41,1	10		376,0	3		61,4	1		276,0	1
249,0	10		37,7	3		227,0	1		26,1	1
37,0	10		22,0	3		348,0	1		76,7	1
50,1	9		83,6	3		220,0	1		93,0	1
17,0	9		38,9	3		65,6	1		81,2	1
14,0	9		53,6	2		1555,0	1		16,9	1
Media										
Brosur	302		Spanduk	301		Internet	280		Rekomendasi	116
Umur						**Rata-Rata Gaji Orang Tua**				
22						2.450.000				

Berdasarkan Tabel 4.22 *Cluster* 2 memiliki karakteristik calon mahasiswa berumur 22 tahun dengan rata-rata gaji orang tua sebesar Rp. 2.450.000. Jarak sekolah ±6,3 Km paling banyak memberikan lulusan dengan jumlah 52 calon mahasiswa dengan persentase 5,2%. Kecamatan yang mempunyai jarak ±5,9 Km mempunyai persentase paling besar dengan 14%. Brosur dan Spanduk sangat berpegaruh dengan perbedaan persentase 0,1%.

Tabel 4.23 Hasil Analisis *Cluster* 3

Cluster 3									
Jarak Asal Sekolah/Km									
18,6	15		6,3	2		6,9	1		103,1	1
35,8	6		16,0	2		18,4	1		251,0	1
38,3	6		14,5	2		13,4	1		2,6	1
17,4	6		4,2	2		250,0	1		70,2	1
9,5	4		28,8	2		10,1	1		1556,0	1
15,3	4		2,5	2		1,7	1		80,6	1
42,4	3		286,0	2		172,0	1		699,0	1
4,6	3		16,3	2		81,4	1		15,2	1
14,2	3		272,0	2		22,4	1		2,0	1
2,1	3		39,6	1		62,8	1		44,3	1
22,2	3		45,1	1		235,0	1		21,0	1
18,0	2		3,2	1		54,3	1		16,9	1
15,5	2		597,0	1		136,0	1			
Jarak Kecamatan/Km										
18,4	31		280,0	3		26,2	1		27,5	1
44,4	13		41,1	2		249,0	1		68,1	1
5,8	9		19,2	2		37,6	1		1555,0	1
35,8	6		9,0	2		1,7	1		53,0	1

Cluster 3										
14,1	4		285,0	2		171,0	1		10,4	1
43,4	3		22,4	2		76,7	1		30,1	1
16,4	3		599,0	1		234,0	1		7,5	1
5,9	3		20,8	1		136,0	1			
22,7	3		14,0	1		251,0	1			
Media										
Spanduk	48		Rekomendasi	25		Internet	21		Brosur	13
Umur						**Rata-Rata Gaji Orang Tua**				
19						4.450.000				

Tabel 4.23 pada *Cluster* 3 memiliki karakteristik calon mahasiswa berumur 19 tahun dengan rata-rata gaji orang tua sebesar Rp. 4.450.000. Jarak sekolah ±18,6 Km paling banyak memberikan lulusan dengan jumlah 15 calon mahasiswa dengan persentase 14%. Kecamatan yang mempunyai jarak ±18,4 Km mempunyai persentase paling besar dengan 29%. Spanduk sangat berpegaruh dengan persentase 44,9%.

Tabel 4.24 Hasil Analisis *Cluster* 4

Cluster 4										
Jarak Sekolah/Km										
42,4	15		35,8	1		2,1	1		315,0	1
11,0	5		326,0	1		1556,0	1		70,2	1
18,7	2		272,0	1		80,6	1		44,0	1
9,5	2		378,1	1		4,2	1		4,6	1
25,0	1		45,0	1		11,1	1		17,4	1
18,6	1		62,8	1		699,0	1		5,0	1
9,6	1		28,8	1		11,6	1			
Jarak Kecamatan/Km										
43,4	20		380,0	1		5,8	1		7,5	1
11,1	3		280,0	1		1,7	1		10,4	1
61,2	3		31,9	1		7,6	1		3,0	1
22,9	3		16,4	1		699,0	1			
18,4	2		1555,0	1		315,0	1			
35,8	1		86,1	1		50,4	1			
Media										
Rekomendasi	20		Spanduk	15		Internet	9		Brosur	3
Umur						**Rata-Rata Gaji Orang Tua**				
19						5.875.532				

Berdasarkan Tabel 4.24 *Cluster* 4 memiliki karakteristik calon mahasiswa berumur 19 tahun dengan rata-rata gaji orang tua sebesar Rp. 5.875.532. Jarak sekolah ±42,4 Km paling banyak memberikan lulusan dengan jumlah 15 calon mahasiswa dengan persentase 31,9%. Kecamatan yang mempunyai jarak ±42,4 Km mempunyai persentase paling besar dengan 31,9%. Rekomendasi sangat berpegaruh dengan perbedaan persentase 42,6%

Berdasarkan Tabel 4.20 sampai Tabel 4.24, *cluster* 0 dapat dikategorikan gaji orang tua sedang, *cluster* 1 dan *cluster* 2 dapat dikategorikan gaji orang tua rendah, sedangkan *cluster* 3 dan *cluster* 4 dapat dikategorikan *cluster* dengan gaji orang tua tinggi. Tabel 4.25 merupakan karakerisik dari setiap *cluster*.

Tabel 4.25 Karakteristik setiap *cluster*

Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Sekolah yang mempunyai jarak ±18,0 paling banyak memberikan lulusan dengan jumlah 29 calon mahasiswa dengan persentase 6%	Sekolah yang mempunyai jarak ±3,2 paling banyak memberikan lulusan dengan jumlah 61 calon mahasiswa dengan persentase 20,4%	Sekolah yang mempunyai jarak ±6,3 paling banyak memberikan lulusan dengan jumlah 52 calon mahasiswa dengan persentase 5,2%	Sekolah yang mempunyai jarak ±18,6 paling banyak memberikan lulusan dengan jumlah 15 calon mahasiswa dengan persentase 14%	Sekolah yang mempunyai jarak ±42,4 paling banyak memberikan lulusan dengan jumlah 15 calon mahasiswa dengan persentase 31,9%
Kecamatan yang mempunyai jarak ±5,8 mempunyai persentase paling besar dengan 39,6%	Kecamatan yang mempunyai jarak ±1,7 mempunyai persentase paling besar dengan 53,2%	Kecamatan yang mempunyai jarak ±5,9 mempunyai persentase paling besar dengan 14%	Kecamatan yang mempunyai jarak ±18,4 mempunyai persentase paling besar dengan 29%	Kecamatan yang mempunyai jarak ±42,4 mempunyai persentase paling besar dengan 31,9%
Brosur dan Rekomendasi sangat berpegaruh dengan perbedaan persentase 3,3%	Brosur sangat berpegaruh dengan persentase 51,5%	Brosur dan Spanduk sangat berpegaruh dengan perbedaan persentase 0,1%	Spanduk sangat berpegaruh dengan persentase 44,9%	Rekomendasi sangat berpegaruh dengan perbedaan persentase 42,6%
Usia paling banyak 19 tahun	Usia paling banyak 23 tahun	Usia paling banyak 22 tahun	Usia paling banyak 19 tahun	Usia paling banyak 19 tahun
Rata-Rata Gaji Orang Tua Rp. 3.450.000	Rata-Rata Gaji Orang Tua Rp. 1.450.000	Rata-Rata Gaji Orang Tua Rp. 2.450.000	Rata-Rata Gaji Orang Tua Rp. 4.450.000	Rata-Rata Gaji Orang Tua Rp. 5.875.532

Berdasarkan hasil pengelompokkan yang terbagi menjadi 5 *cluster*, dapat dilakukan strategi promosi berdasarkan *promotion mix* pada Tabel 4.26 menggunakan atribut-atribut yang saling berkaitan dan dari hasil wawancara dengan bagian PUSINFO UNISA.

Tabel 4.26 Strategi Promosi Berdasarkan *Promotion Mix*

No	Strategi Promosi	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	**Periklanan**					
	Brosur	√	√	√		
	Spanduk			√	√	
2	**Penjualan Personal**					
	Rekomendasi	√				√
	Persentasi Sekolah	√	√	√		
3	**Promosi Penjualan**					
	Voucher Diskon	√	√	√		
4	**Hubungan Masyarakat**					
	Membuka Stand di Kecamatan	√	√	√		
	Mengadakan Perlombaan	√	√	√	√	√
	Kerjasama Sekolah	√	√	√	√	√
5	**Pemasaran Langsung**					
	Internet	√	√	√	√	√

Rekomendasi promosi untuk UNISA Kuningan berdasarkan *Promotion Mix* terlampir (Lampiran 4).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan yang sudah dilakukan, diperoleh kesimpulan-kesimpulan sebagai berikut:

1. Nilai K optimum pada algoritma K-Means berada pada K=5 dengan nilai dunn index paling tinggi 0,999, sedangkan nilai K optimum pada K-Medoids berada pada K=4 dengan nilai dunn index paling tinggi 0,499.
2. Algoritma K-Means menjadi algoritma yang lebih baik dibandingkan dengan K-Medoids. Hal ini dapat dilihat berdasarkan dari nilai validasi dunn Index paling tinggi pada K-Means K=5 sebesar 0,999 dan dalam proses pengolahannya K-Means hanya membutuhkan waktu rata-rata 1 detik, sedangkan algoritma *K-Medoids* membutuhkan waktu rata-rata 1 menit yang artinya apabila jumlah pengelompokkan yang ditentukan lebih besar, proses pengolahan datapun semakin lama.
3. Profil calon mahasiswa yang dihasilkan algoritma *K-Means* dapat memberikan pengetahuan berupa strategi promosi berdasarkan *promotion* mix, dimana brosur, persentasi sekolah,dan pemberian *voucher* diskon lebih diutamakan pada *cluster* 0, *cluster* 1 dan *cluster* 2, spanduk lebih diutamakan pada *cluster* 2 dan *cluster* 3, dan rekomendasi lebih diutamakan pada *cluster* 0 dan *cluster* 4. Membuka stand di kecamatan lebih diutamakan pada *cluster* 0, *cluster* 1, dan *cluster* 2. Mengadakan perlombaan, kerjasama sekolah dan internet dapat dilakukan pada semua *cluster.*

5.2 Saran

Adapun saran dari penelitian diantaranya :

1. Pada penelitian berikutnya diharapkan agar bisa membandingkan dengan menggunakan algoritma *cluster*ing lainnya.
2. Menambahkan atribut lain yang tidak digunakan pada penelitian ini yang dapat menghasilkan strategi promosi yang lebih baik.

DAFAR PUSTAKA

Chasanah, T. T. (2017). Penentuan Strategi Promosi Penerimaan Mahasiswa Baru dengan Algoritma Clustering K-Means. *Integrated Circuit Technology (IC-TECH)*, Vol. 12 No. 1.

Chowdary, N. S. (2014). Evaluating and Analyzing Clusters in Data Mining using Different Algorithms. *International Journal of Computer Science and Mobile Computing*, Vol. 3 No. 2.

Dewi, I. C. (2017). Analysis of Clustering for Grouping of Productive Industry by K-Medoid Method. *International Journal of Engineering and Emerging Technology*, Vol. 2 No. 1.

Gandhi, G. (2014). Analysis and Implementation of Modified K-Medoids Algorithmn to Increase Scalibility and Efficiency for Large Dataset. *International Journal of Research in Engineering and Technology (IJRET)*, Vol. 3 No. 6.

Guoli, L. d. (2013). The Improved Research on K-Means Clustering Algorithm in Initial Values. *International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*.

Hermanto, D. M. (2017). Analisis Algortma Clustering dalam Kasus Penentuan Jenis Bunga IRIS. *JURNAL MEDIA APLIKOM.*, Vol. 9 No. 2.

Irwansyah, E. (2017, 03 09). *Clustering*. Diambil kembali dari School of Computer Science BINUS UNIVERSITY: https://socs.binus.ac.id/2017/03/09/clustering/

Klinkenberg, H. M. (2013). *RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series),".* U.S.A: CRC Press.

Kurniawati, I. (2017). Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru. *IKRAITH-INFORMATIKA*, Vol. 1 No. 2.

Larose, D. T. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining.* Canada: WILEY.

Liu, Y. (2010). Understanding of Internal Clustering Validation Measures. *IEEE International Conference on Data Mining*.

Maimon, L. R. (2010). *Data Mining and Knowledge Discovery Handbook, 2nd ed.* London: Springer Science+Business Media.

Marlina, D. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*, Vol.4, No.2.

Monalisa, S. (2018). KLASTERISASI CUSTOMER LIFETIME VALUE DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA K-MEANS . *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 5, No. 2.

Nanda, A. (2016). Comparative Study between Parallel K-Means and Parallel K-Medoids using Message Passing Interface (MPI). *International Journal on Information and Communication of Technology*, Vol. 2 No. 2.

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Pengembangan IT (JPIT)*, Vol. 4 No. 1.

Pramesti, D. F. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 1 No. 9.

Rendón, E. d. (2011). Internal Versus External Cluster Validation. *INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS*, Vol. 5 No. 1.

Sibarani, R. (2018). Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia [Algorithma K-Means Clustering Strategy Marketing Admission Universias Satya Negara Indonesia]. *Seminar Nasional Cendekiawan*, Vol. 1 No.2.

Sibuea, F. L. (2017). Pemetaan Siswa Berpresasi menggunakanMeode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi*, Vol. 4 No. 1.

Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data.* Bandung: INFORMATIKA.

Syarfan, A. d. (2016). ANALISIS BAURAN PROMOSI (PROMOTION MIX) PRODUK MULTILINKED SYARIAH PADA ASURANSI PANIN DAI-ICHI LIFE CABANG PEKANBARU. *Jurnal Valuta*, Vol. 2 No. 1.

Talakua, M. W. (2017). Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Jurnal Ilmu Matematika dan Terapan*, Vol. 11 No. 2.

Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making.* Italy: WILEY.

Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer.* Yogyakarta: Gava Media.

Wang, K. (2009). Validation for Cluster Analyses. *Data Science Journal*, Vol. 8.

Widiyaningtyas, T. d. (2017). Implementation of K-Means Clustering Method to Distribution of High School Teachers. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*.

Yunita. (2018). Penerapan data mining menggunakan algoritma K-Means Clusering pada penerimaan mahasiswa baru (Studi kasus : Universitas Islam Indragiri). *Jurnal SISTEMASI*, Vol. 7, No. 3.