KAJIAN DATA *MINING* DENGAN DATA CALON SISWAUNTUK MENENTUKAN STRATEGI PROMOSI SEKOLAH MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS*

(Studi Kasus: SMK Bhakti Putra)

TESIS

Disusun Sebagai Salah Satu Syarat Untuk
Memperoleh Gelar Magister Komputer
Dari Sekolah Tinggi Manajemen Informatika Dan Komputer LIKMI

Oleh:

UJANG SARIFUDIN

NPM : 2019210089

PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2021

KAJIAN DATA *MINING* DENGAN DATA CALON SISWA UNTUK MENENTUKAN STRATEGI PROMOSI SEKOLAH MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS*

(Studi Kasus: SMK Bhakti Putra)

Oleh :

UJANG SARIFUDIN

NPM : 2019210089

Bandung, 12 April 2021

Menyetujui,

Dr. Djajasukma Tjahjadi, S.E., M.T.

Pembimbing

PROGRAM STUDI PASCASARJANA

MAGISTER SISTEM INFORMASI

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI

BANDUNG

2021

ABSTRAK

KAJIAN DATA *MINING* DENGAN DATA CALON SISWA UNTUK MENENTUKAN STRATEGI PROMOSI SEKOLAH MENGGUNAKAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* (Studi Kasus: SMK Bhakti Putra)

Ujang Sarifudin
NPM 2019210089

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 Sekolah Menengah Kejuruan. Penerapan teknologi pada Sekolah dapat diperoleh dari data historis, data akan terus bertambah sehingga terjadi penumpukan data dan dapat memperlambat pencarian informasi terhadap data tersebut.

SMK Bhakti Putra merupakan sekolah swasta di Kabupaten Bandung Jawa Barat yang harus berusaha dalam meningkatkan jumlah penerimaan siswa baru pada setiap tahun ajaran baru, karena jumlah siswa dapat mempengaruhi kegiatan pembelajaran, operasional dan pengembangan bagi Sekolah selain itu didasari juga jumlah siswa yang tidak stabil di tiap tahun nya, terkadang bisa naik ataupun turun jumlah siswa nya serta belum adanya strategi promosi juga menjadi masalah untuk bisa dipecahkan.

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*, Algoritma K-Means clustering merupakan algoritma yang berperan penting dalam bidang data mining serta sederhana untuk diimplementasikan dan dijalankan. Pengolahan terhadap data yang terus menumpuk dapat di lakukan untuk mengetahui informasi yang tersembunyi, Pada proses Klasifikasi ini Data mining akan memiliki manfaat yaitu mendeskripsikan data yang penting dan juga dapat meramalkan kecenderungan data pada masa depan.

Hasil dari penelitian ini yaitu rekomendasi strategi promosi untuk sekolah yang diteliti didapat 5 *cluster* setelah dianalisis Profil calon siswa yang dihasilkan algoritma *K-Means* dapat memberikan pengetahuan berupa strategi promosi berdasarkan *promotion mix*, dimana brosur, persentasi sekolah,dan pemberian beasiswa lebih di utamakan pada *cluster* 0, *cluster* 1 dan *cluster* 3, Spanduk lebih di utamakan pada *cluster* 0 dan *cluster* 2, dan rekomendasi lebih di utamakan pada *cluster* 2 dan *cluster* 4. Membuka stand di kecamatan lebih diutamakan pada *cluster* 0, *cluster* 1, *cluster* 2 dan *cluster* 3. Mengadakan perlombaan dan kerjasama dapat dilakukan pada semua *cluster*.

Kata Kunci : Data *Mining*, *K-Means, K Medoids*. *Promotion Mix, Dunn Index*

ABSTRACT

STUDY DATA MINING WITH PROSPECTIVE STUDENTS TO DETERMINE A SCHOOL PROMOTION STRATEGY USING K-MEANS AND K-MEDOIDS ALGORITHM
(*Case Study*: SMK Bhakti Putra)

Ujang Sarifudin
NPM 2019210089

The rapid progress and development of information technology can facilitate the process of searching for information and in producing information such as in the fields of economy, industry, technology, including the field of Vocational High School education. The application of technology in schools can be obtained from historical data, the data will continue to increase, resulting in data accumulation and can slow down the search for information on that data.

Bhakti Putra School is a private school in Bandung Regency, West Java which must try to increase the number of new student admissions every new academic year, because the number of students can affect the learning, operational and development activities for the school. years, sometimes the number of students can increase or decrease and the absence of a promotion strategy is also a problem to be solved.

Data mining is an activity in extracting hidden information from the database by extracting patterns from the data, the results of which can be used as further decision making. One of the data mining techniques is clustering, which is the process of dividing data into groups based on predetermined characteristics. There are many algorithms that can be used for clustering, one of which is the K-Means clustering algorithm and K-Medoids clustering, the K-Means clustering algorithm is an algorithm that plays an important role in the field of data mining and is simple to implement and run. Processing of data that continues to accumulate can be done to find out hidden information. In this classification process data mining will have the benefit of describing important data and can also predict data trends in the future.

The results of this study are the recommendations for promotion strategies for the school under study obtained 5 clusters after analyzing the profiles of prospective students produced by the K-Means algorithm can provide knowledge in the form of promotion strategies based on the promotion mix, where brochures, school percentages, and scholarships are prioritized on the cluster 0, cluster 1 and cluster 3, Banners are prioritized for cluster 0 and cluster 2, and recommendations are prioritized for cluster 2 and cluster 4. Opening stands in sub-districts is preferred to cluster 0, cluster 1, cluster 2 and cluster 3. Organizing competitions and collaborations can be done in all clusters.

Keywords: Data Mining, K-Means, K Medoids.Promotion Mix, Dunn Index

KATA PENGANTAR

Puji serta syukur penulis panjatkan ke hadirat Allah SWT, berkat rahmat, karunia serta pertolongan-Nya penulis dapat menyelesaikan Tesis ini dengan lancar. Shalawat serta salam penulis limpah curahkan kepada Nabi besar Muhammad Rosulullah SAW sebagai suri tauladan bagi kita semua.

Penulisan Tesis ini bertujuan sebagai salah satu syarat untuk memperoleh gelar Magister Komputer dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI.

Dengan tidak mengurangi makna syukur dan dengan segala ketulusan serta kerendahan hati penulis mengucapkan banyak terima kasih kepada :

1. Bapak Dr. Djajasukma Tjahjadi, S.E., M.T., selaku pembimbing dalam penulisan Tesis ini, yang telah membimbing dan memberikan pengarahan kepada penulis.
2. Bapak Bagus Abdurrahman, S.Kom., selaku ketua PSB, yang telah membantu dalam memberikan kesempatan untuk melakukan penelitian.
3. Orang tua, Istri, Anak - anaku serta yang keluarga tercinta atas dukungannya, baik secara moril maupun materil.
4. Rekan-rekan STMIK LIKMI khususnya kelas S tahun angkatan 2019.
5. Kepada semua pihak yang telah membantu penyelesaian Tesis ini.

Penulis menyadari dalam penulisan ini masih banyak sekali kekurangan karena keterbatasan pengetahuan dan pengalaman dari penulis, saran dan kritik yang bersifat membangun sangat penulis harapkan untuk kesempurnaan Tesis ini di masa yang akan datang. Semoga ketulusan dan kebaikan yang telah diberikan kepada penulis mendapatkan balasan dari Allah SWT. Amin

Bandung, April 2021

Penulis

DAFTAR ISI

ABSTRAK i

ABSTRACT ii

KATA PENGANTAR iii

DAFTAR ISI iv

DAFTAR GAMBAR vi

DAFTAR TABEL vii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 3

1.3 Tujuan Penelitian 3

1.4 Batasan Masalah 4

1.5 Sistematika Penulisan 4

BAB II LANDASAN TEORI 5

2.1 Data *Mining* 5

2.1.1 *Preprocessing Data* 6

2.1.2 *Clustering* 7

2.1.3 Algoritma *K-Means* 9

2.1.4 Algoritma *K-Medoids* 11

2.2 *Dunn Index* 13

2.3 Rapidminer 13

2.4 Promosi 14

2.5 Penelitian Terkait 17

BAB III OBJEK DAN METODOLOGI PENELITIAN 19

3.1 Profil SMK Bhakti Putra 19

3.2 Metodologi Penelitian 20

BAB IV HASIL DAN PEMBAHASAN 28

4.1 *Selection* Data 28

4.2 *Preprocessing* Data 28
4.3 *Transformasi* Data 31
4.4 *Implementasi Algoritma Clustering* 33
4.5 *Pattern Evaluation* 38
4.6 *Interpretation Knowledge* 40
4.7 Strategi Promosi 48
BAB V KESIMPULAN DAN SARAN 58
5.1 Kesimpulan 58
5.2 Saran 59
DAFTAR PUSTAKA 60
DAFTAR LAMPIRAN

DAFTAR GAMBAR

Gambar 1.1 Grafik Calon Siswa Baru Tahun 2017-2020 1

Gambar 2.1 Tahapan KDD 5

Gambar 2.2 Merubah data kosong dengan konstanta yang ditentukan analisis 7

Gambar 2.3 Merubah data kosong dengan mean dan modus 7

Gambar 2.4 *Flowchart* Algoritma *K-Means* 10

Gambar 2.5 *Flowchart* Algoritma *K-Medoids* 12

Gambar 3.1 Metodologi Penelitian Berdasarkan KDD 22

Gambar 4.1 Penambahan Operator *Read Excel* dan *Nominal to Numerical* 32

Gambar 4.2 Penambahan Operator *Multiply* dan Algoritma 33

Gambar 4.3 *Cluster* model K = 2 pada *K-Means* (a) dan *K-Medoids* (b)..... 34

Gambar 4.4 *Cluster* model K = 3 pada *K-Means* (a) dan *K-Medoids* (b)..... 35

Gambar 4.5 *Cluster* model K = 4 pada *K-Means* (a) dan *K-Medoids* (b)..... 35

Gambar 4.6 *Cluster* model K = 5 pada *K-Means* (a) dan *K-Medoids* (b)..... 36

Gambar 4.7 *Cluster* model K = 6 pada *K-Means* (a) dan *K-Medoids* (b)..... 36

Gambar 4.8 *Cluster* model K = 7 pada *K-Means* (a) dan *K-Medoids* (b)..... 37

Gambar 4.9 Penambahan operator validasi dunn *Index* 38

Gambar 4.10 Grafik *Cluster K-Means* dan *K-Medoids* terhadap Nilai Dunn *Index* 39

Gambar 4.11 *Cluster Model* Algoritma *K-Means* pada K = 5 40

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait 17

Tabel 2.2 Perbandingan Penelitian 18

Tabel 3.1 Sepuluh atribut pada data siswa baru SMK Bhakti Putra 24

Tabel 3.2 Data Kosong Pada Setiap Atribut Tahun 2017-2020 25

Tabel 3.3 Pengisian Data Kosong dengan Modus 25

Tabel 4.1 Pemilihan Atribut yang digunakan 27

Tabel 4.2 Pemilihan Atribut yang dihilangkan 27

Tabel 4.3 Pengisian Data Kosong dengan Modus 28

Tabel 4.4 Pengisian Data Kosong dengan Mean 28

Tabel 4.5 Pengisian Data Kosong dengan *Random* 29

Tabel 4.6 Perbaikan Data Tidak Konsisten 29

Tabel 4.7 Perbaikan Data Dalam Pengetikan Data 29

Tabel 4.8 Menghilangkan Data Yang Mempunyai Banyak Data Kosong 30

Tabel 4.9 Hasil Integrasi Tabel Tahun 2017 sampai 2020 30

Tabel 4.10 Transformasi Alamat Menjadi Kecamatan 31

Tabel 4.11 Transformasi Gaji Orang Tua Menggunakan *Means* 31

Tabel 4.12 Transformasi Data Atribut Jurusan 32

Tabel 4.13 Data Hasil Transformasi Pada Data Siswa Baru 33

Tabel 4.14 Parameter *Clustering K-Means* 34

Tabel 4.15 Parameter *Clustering K-Medoids* 34

Tabel 4.16 *Cluster K-Means* dan *K-Medoids* terhadap Nilai Dunn *Index* 38

Tabel 4.17 Hasil Perhitungan Jarak *Cluster* dengan *Centroid* 38

Tabel 4.18 Hasil Analisi *Cluster* 0 41

Tabel 4.19 Hasil Analisis *Cluster* 1 42

Tabel 4.20 Hasil Analisis *Cluster* 2 43

Tabel 4.21 Hasil Analisis *Cluster* 3 45

Tabel 4.22 Hasil Analisis *Cluster* 4 46

Tabel 4.23 Karakteristik setiap *cluster* 48

Tabel 4.24 Strategi Promosi Berdasarkan *Promotion Mix* 48

Tabel 4.25 Nama Asal Sekolah untuk Brosur 51

Tabel 4.26 Jurusan Banyak Diminati Pada Media Brosur 51

Tabel 4.27 Jurusan Kurang Diminati Pada Media Brosur..... 52

Tabel 4.28 Nama Kecamatan Untuk Media Spanduk..... 52

Tabel 4.29 Nama Program Studi Banyak Diminati pada Media Spanduk 52

Tabel 4.30 Nama Program Studi Kurang Diminati pada Media Spanduk 53

Tabel 4.31 Nama Sekolah untuk Rekomendasi 53

Tabel 4.32 Tabel Nama Program Studi Banyak Diminati pada Media Rekomendasi 53

Tabel 4.33 Tabel Nama Program Studi Kurang Diminati pada Media Rekomendasi..... 54

Tabel 4.34 Kecamatan Berdasarkan Rata-Rata Gaji Menengah dan Rendah..... 55

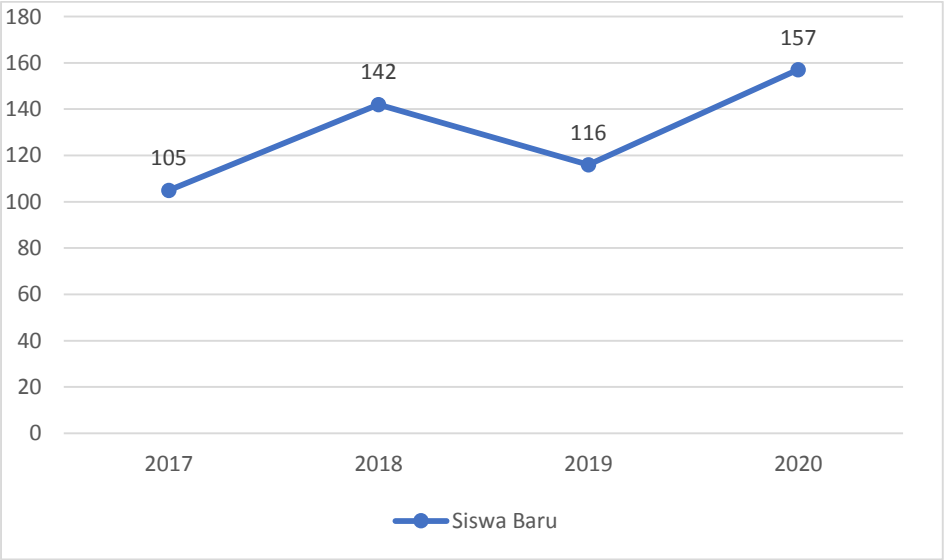
BAB I
PENDAHULUAN

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 Sekolah Menengah Kejuruan. Penerapan teknologi pada Sekolah dapat diperoleh dari data historis, data akan terus bertambah sehingga terjadi penumpukan data dan dapat memperlambat pencarian informasi terhadap data tersebut.

SMK Bhakti Putra merupakan sekolah swasta di Kabupaten Bandung Jawa Barat yang harus berusaha dalam meningkatkan jumlah penerimaan siswa baru pada setiap tahun ajaran baru, karena jumlah siswa dapat mempengaruhi kegiatan pembelajaran, operasional dan pengembangan bagi Sekolah selain itu didasari juga jumlah siswa yang tidak stabil di tiap tahun nya, terkadang bisa naik ataupun turun jumlah siswa nya serta belum adanya strategi promosi juga menjadi masalah untuk bisa dipecahkan.

Setiap tahun SMK Bhakti Putra melakukan penerimaan siswa baru. Jumlah siswa baru di SMK Bhakti Putra dari tahun 2017 sampai tahun 2020 seperti pada Gambar 1.1.



Gambar 1.1

Grafik Calon Siswa Baru Tahun 2017-2020

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*, Algoritma K-Means clustering merupakan algoritma yang berperan penting dalam bidang data *mining* serta sederhana untuk diimplementasikan dan dijalankan. Pengolahan terhadap data yang terus menumpuk dapat di lakukan untuk mengetahui informasi yang tersembunyi, Pada proses Klasifikasi ini data *mining* akan memiliki manfaat yaitu mendeskripsikan data yang penting dan juga dapat meramalkan kecenderungan data pada masa depan.

Dalam prakteknya, algoritma ini umum digunakan karena relatif cepat dan mudah beradaptasi. Bertujuan mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang terdapat dalam dataset, penggunan medoids tidak berdasarkan dari pengamatan meanyang dimiliki oleh setiap *cluster*. Algoritma *K-Medoids* muncul sebagai penanggulangan kelemahan Algoritma *K-Means* yang sensitif terhadap *outlier* dikarenakan sebuah objek dengan sebuah nilai yang besar mungkin secara substansial menyimpang dari distribusi data.

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. 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), 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 Siswa baru tahun 2017 sampai tahun 2020 di SMK Bhakti Putra 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 hasil perbandingan pengolahan data menggunakan Algoritma *K-Means* dan *K-Medoids* ?
2. Bagaimana strategi promosi yang tepat untuk setiap cluster calon siswa di SMK Bhakti Putra ?

1.3 Tujuan Penelitian

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

1. Mengetahui hasil perbandingan pengklasteran menggunakan metode *K-Means* dan *K-Medoids*.
2. Mengetahui strategi promosi yang tepat untuk setiap cluster calon siswa di SMK Bhakti Putra.

1.4 Batasan Masalah

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

1 Data yang digunakan yaitu data Siswa Baru dari tahun 2017 sampai tahun 2020 di SMK Bhakti Putra.

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

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 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 paling 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 dengan data calon siswa.

Bab V. Kesimpulan dan Saran

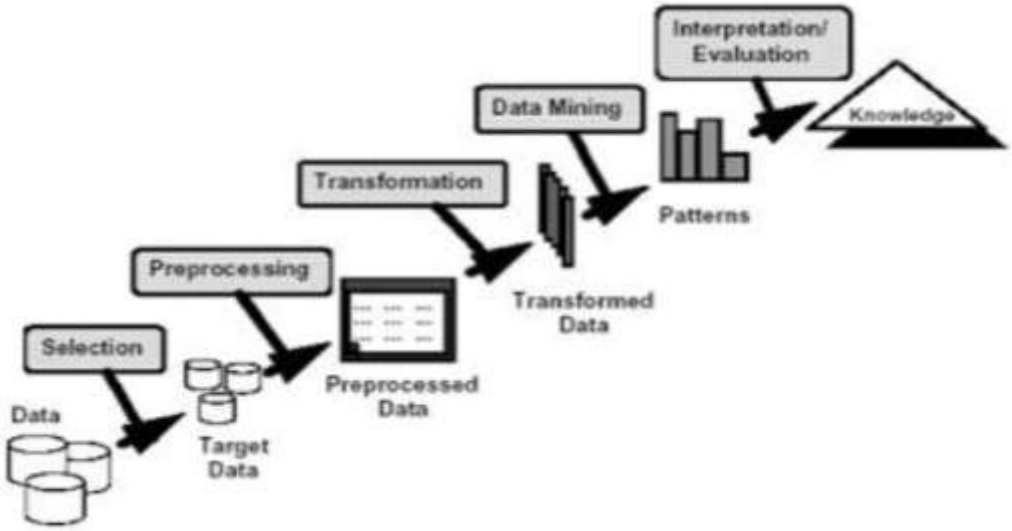
Bab ini menjelaskan tentang kesimpulan penelitian serta saran-saran yang dapat diberikan untuk penelitian selanjutnya.

BAB II
LANDASAN TEORI

2.1 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.

2.1.1 *Preprocessing Data*

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

1. Bidang yang sudah usang (kadaluwarsa)
2. Nilai yang hilang (*missing values*)
3. Pencilan (*outlier)*

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 hilang dengan konstanta, ditentukan oleh analis. Contoh pada Gambar 2.2.

	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	0	71	Europe
3	17.000	302	140	US
4	15.000	400	150	Missing
5	37.700	89	62	Japan

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 diisi dengan nilai yang paling banyak keluar.

2.1.2 *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 *Index validitas* untuk memilih *Number of Cluster* (NC). *Index validitas* digunakan untuk mengukur kualitas hasil *clustering*. Ada dua jenis *Index validitas*, yaitu *Index eksternal* dan *Index internal*. *Index eksternal* adalah ukuran perjanjian antara dua partisi di mana partisi pertama adalah struktur pengelompokan apriori, dan hasil kedua dari prosedur. Index internal digunakan untuk mengukur kebaikan struktur pengelompokan tanpa menggunakan informasi *eksternal*. Untuk *Index eksternal*, mengevaluasi hasil dari

algoritma pengelompokan berdasarkan struktur *cluster* yang diketahui dari kumpulan data (label *cluster*), sedangkan Index internal mengevaluasi hasilnya menggunakan jumlah dan fitur yang melekat dalam kumpulan data. NC yang optimal biasanya ditentukan berdasarkan internal *Index*.

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 *Index* validasi *internal* yang sesuai dari setiap partisi yang diperoleh pada Langkah 2.
4. Pilih partisi terbaik dan jumlah *cluster* optimal sesuai dengan kriteria.

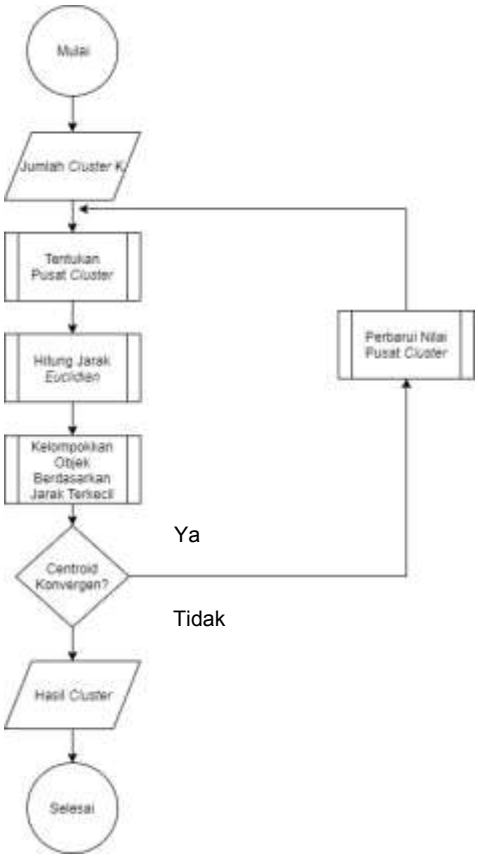
2.1.3 Algoritma *K-Means*

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 (Guoil, 2013).

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 2.4:



Gambar 2.4
Flowchart Algoritma *K-Means*

Berdasarkan Gambar 2.4 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) :

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

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.

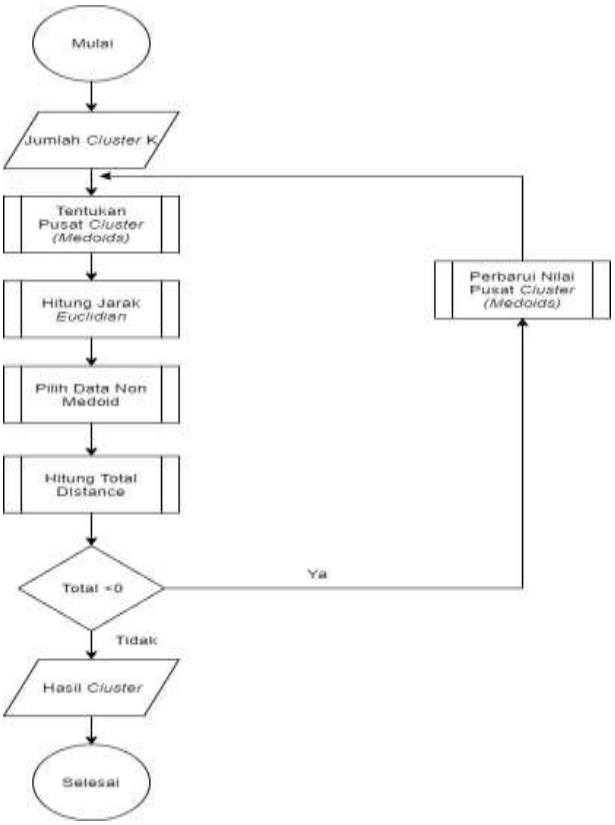
2.1.4 Algoritma *K-Medoids*

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 2.5
Flowchart Algoritma *K-Medoids*

Berdasarkan Gambar 2.5 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. Masukkan 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.

2.2 Dunn *Index*

Dunn Index merupakan salah satu Index validasi internal, dimana Index 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 Index adalah dengan menghitung nilai minimum dari perbandingan antara nilai fungsi disimilaritas 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 Index (Chowdary N. S., 2014):

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

Dimana :

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

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

2.3 *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 *rapidminer* 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).

2.4 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)

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.

A. Data Mining dan Promosi Sekolah

Promosi sekolah berarti suatu cara untuk menyampaikan informasi kepada masyarakat tentang keberadaan sekolah dengan tujuan menarik calon siswa agar mereka mendaftar dan bersekolah di sekolah yang kita promosikan. Sederhana nya sepeti itu Intinya bagaimana cara mendapatkan siswa baru. Kalaupun ada tujuan lainnya seperti meningkatkan reputasi sekolah dan pencapaian lainnya, memperkenalkan iklim sekolah serta keunggulan khas sekolah dan lain sebagainya. Pada akhirnya akan mengerucut pada peningkatan perolehan siswa baru, Cara promosi yang paling kuno dan ternyata paling epektif adalah promosi dari mulut ke mulut, atau dalam bahasa marketing dikenal dengan istilah *word of mouth* (WOM). Promosi semacam ini lahir dari kepuasan konsumen. Kalau orang tua siswa merasa puas dengan program sekolah, kualitas guru, fasilitas dan prestasi-prestasi yang diperoleh sekolah maka dengan sendirinya para orang tuas siswa tersebut dengan suka rela membicarakannya ke saudara, tetangga atau kenalan barunya.

Ini menyangkut reputasi dan pelayanan sekolah tentunya. Makanya sekolah dengan kualitas diatas rata-rata akan mempunyai relawan-relawan promosi yang banyak karena setiap orang tua siswa akan menceritakan mempromosikan sekolah tempat anaknya bersekolah terus menerus tidak hanya ketika waktu Penerimaan siswa baru.

Data *Mining* digunakan untuk menjabarkan penemuan ilmu pengetahuan di dalam *database*. Dalam pemanfaatan data *mining* di bidang pendidikan khususnya promosi memeberikan manfaat agar promosi-promosi yang telah dilakukan selama ini bisa lebih efisien dan efektif. Untuk dapat melakukan promosi-promosi yang lebih efektif dan efisien maka peneliti akan mengolah semua data yang ada sehingga menghasilkan strategi promosi yang lebih baik lagi. Pengolahan data yang dilakukan adalah dengan cara pengelompokan menggunakan algoritma *K-Medoids* terhadap data para calon siswa antara lain Nama, Tempat dan Tanggal Lahir, Alamat, Agama, Nomor Telepon, Asal Sekolah, Program Studi Pilihan dan Pembawa Informasi, yang mana data ini diperoleh dari kegiatan promosi yang dilakukan dari tahun 2019 hingga 2020 (Anggraeni, 2019).

Siswa-siswi SMK Bhakti Putra berasal dari kecamatan nagreg maupun luar kecamatan nagreg maka dibutuhkan strategi khusus dan beberapa cara dalam melakukan pemasaran untuk mencari calon siswa-siswi agar promosi yang dilakukan lebih efektif dan efisien. penyebaran pamflet yang disebarkan disembarang tempat dan promosi tidak terorganisir maka menyebabkan tidak efisiennya untuk pemasaran sering kami timbul ketidak pastian data. Untuk menjawab semua pertanyaan tersebut maka diadakan sebuah penelitian yang dilakukan dengan cara mengolah data-data siswa seperti nama siswa, daerah asal, jurusan yang diambil serta yang terakhir nilai akademik siswa yang merupakan salah satu faktor literatur penting dalam dunia Pendidikan (Surmaningtyas, 2015).

Metode pengolahan data seperti ini sering disebut sebagai data mining. Serta algoritma ini merupakan algoritma yang paling banyak digunakan karena mempunyai kemampuan dalam mengelompokkan data dengan waktu komputasi yang cepat, Efektif dan efisien dengan menggunakan jumlah data yang cukup besar Ide dasar algoritma *nya* sangatlah sederhana, yaitu meminimalkan *Sum of Squared Error* (SSE) antara objek-objek data.

2.5 Penelitian Terkait

Berikut ini merupakan beberapa penelitian yang terkait dengan *clustering* menggunakan *K-Means dan K Medoids* 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* 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	Choirul,2018	Penelitian ini melakukan pengelompokkan calon data siswa untuk menentukan ada berapa cluster untuk menentukan mengelompokkan siswa supaya dapat menentukan strategi promosi sekolah Penelitian ini diharapkan dapat dijadikan sebagai referensi untuk penelitian selanjunya dan melakukan penerapan dengan metode *clustering* yang dipakai.

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, pada penelitian yang 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 Index karena berdasarkan Tabel 2.2 belum ada penelitian yang membandingkan Algoritma *K-Means* dan Algoritma *K-Medoids* menggunakan Dunn Index.

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	√	-	-	-
Choirul, 2018	√	-	-	-

BAB III
OBJEK DAN METODOLOGI PENELITIAN

3.1 Profil SMK Bhakti Putra

Penelitian berorientasi pada strategi promosi di SMK Bhakti Putra Nagreg Kab. Bandung yang merupakan sekolah menengah kejuruan swasta yang terletak di jl. Andir No 98 Nagreg Kabupaten Bandung Provinsi Jawa Barat yang telah resmi dan diakui oleh Kementrian Pendidikan dan Kebudayaan Republik Indonesia. Pada tanggal 12 Januari 2017, Yayasan Pendidikan Bhakti Putra dengan didukung Pemerintah Daerah Kabupaten Bandung prakarsa dalam mendirikan SMK Bhakti Putra. Tahun 2017 dengan Keputusan Kordinator Dinas Pendidikan Provinsi Jawa Barat nomor : 1005 tahun 2017 diperoleh ijin prinsip operasionalnya. Status ijin operasional berjalan dari tahun 2017 sampai Sekarang. SMK Bhakti Putra memiliki dua kompetensi keahlian yaitu Rekayasa Perangkat Lunak Dan Bisnis Daring Dan Pemasaran. Berikut adalah Visi Misi SMK Bhakti Putra :

VISI

“Mewujudkan Sumber daya manusia yang berkualitas dalam keimanan dan dibarengi dengan ilmu pengetahuan dan ilmu teknologi”.

MISI

1. Menyelenggarakan Pendidikan dengan standarisasi Pendidikan nasional berbasis teknologi Informatika.
2. Menyiapkan calon pemimpin masa depan yang memiliki ilmu pengetahuan dan mempunyai ilmu pengetahuan dan memiliki iman dan taqwa yang kuat.
3. Pengembangan potensi siswa berbasis talenta, dengan meniti beratkan pada seni, Olahraga dan jiwa berwirausaha
4. Pengembangan suasana belajar ramah lingkungan, dengan penataan ruangan belajar dan lingkungan sekolah hijau *(Green School).*
5. Memotivasi peserta didik untuk belajar menerapkan kemajuan teknologi.

6. Menyelenggarakan Pendidikan berstandar Nasional yang mengedepankan keterampilan akademik dan umum

Aktivitas promosi mulai dari sosialisasi ke SMP/MTS di kecamatan nagreg dan kecamatan - kecamatan sekitarnya adapun kegiatan yang dilaksanakan didalamnya adalah penyebaran angket minat siswa salah satu tujuan penyebaran angket minat siswa adalah seberapa besar minat siswa SMP/MTS melanjutkan *study,* dari penyebaran angket tersebut dikemas dalam hasil survey angket minat siswa.

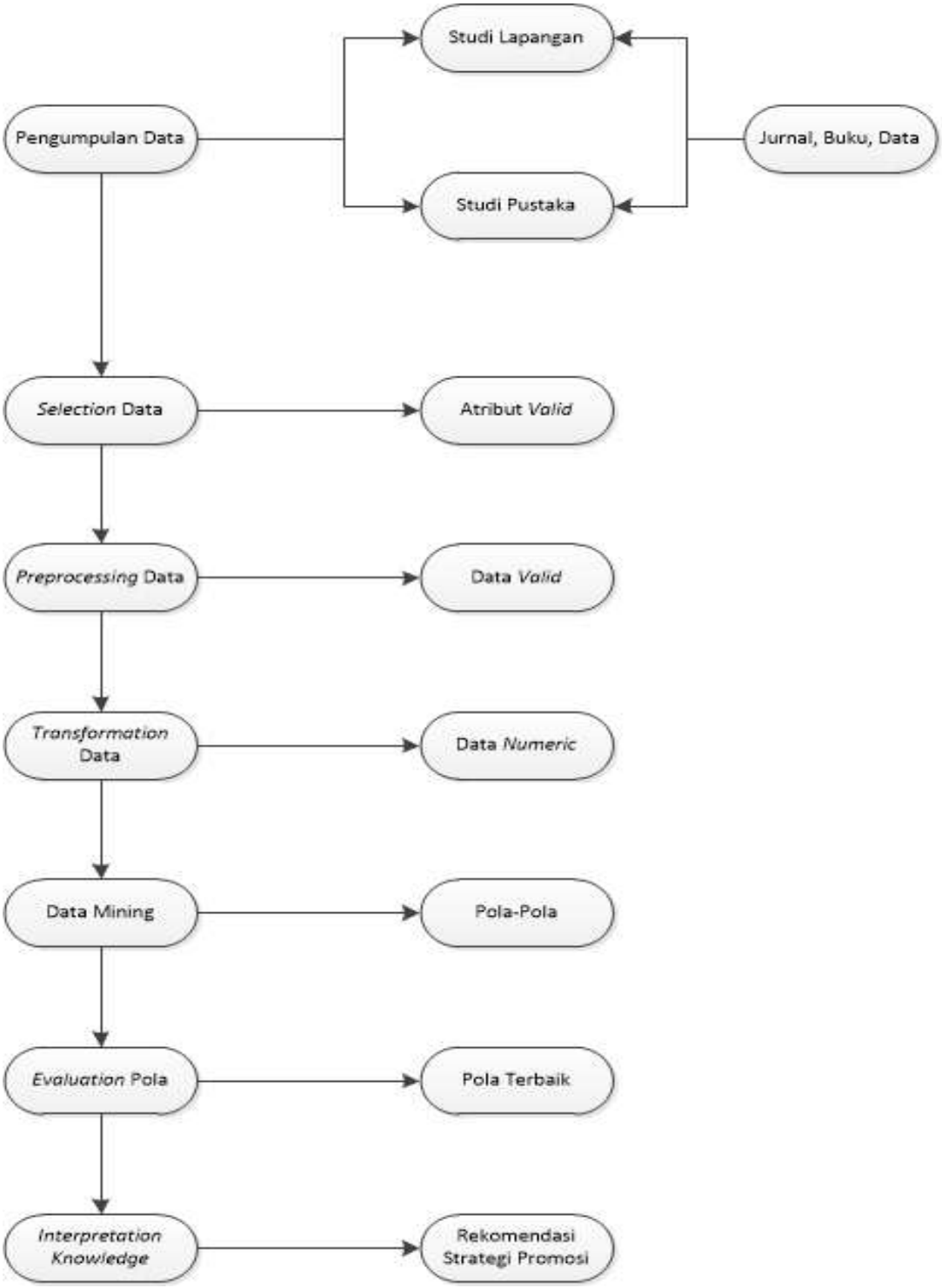
3.2 Metodologi Penelitian

Knowledge Discovery and Data Mining (KDD) adalah proses yang dibantu oleh komputer untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi dan pengetahuan yang berguna. Data mining *tools* memperkirakan perilaku dan tren masa depan, memungkinkan bisnis untuk membuat keputusan yang proaktif dan berdasarkan pengetahuan. Data mining tools mampu menjawab permasalahan bisnis yang secara tradisional terlalu lama untuk diselesaikan. Data mining tools menjelajah database untuk mencari pola tersembunyi, menemukan infomasi yang prediktif yang mungkin dilewatkan para pakar karena berada di luar ekspektasi mereka.

KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interprestasi dan visualisasi dari pola-pola sejumlah kumpulan data. *Knowledge discovery in databases* (KDD) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Banyak perusahaan telah mengumpulkan data berkuantitas besar. Teknik data *mining* dapat diimplementasikan ke *platform software* dan hardware yang sudah ada untuk meningkatkan nilai dari sumberdaya informasi yang sudah ada. Implementasi data *mining* pada *client/server* berperforma tinggi atau komputasi paralel dapat menganalisis *database* yang besar untuk menjawab pertanyaan seperti, Klien mana yang paling mungkin merespon surat promosi kami selanjutnya, dan kenapa ?

KDD merupakan metodologi yang digunakan pada penelitian ini

Gambar 3.1 merupakan tahapan-tahapan penelitian yang dilakukan pada penelitian ini.



Gambar 3.1
Metodologi Penelitian Berdasarkan KDD

Dari Gambar 3.1, 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 Siswa. Observasi digunakan untuk melihat keadaan nyata di SMK Bhakti Putra yang menjadi objek penelitian, wawancara untuk mengetahui strategi promosi apa saja yang pernah dilakukan, adapun istilah yang muncul dan dipelajari disini sepeti Personal Selling, Rekomendasi dan Media promosi untuk memperoleh data Calon siswa 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 Studi kepustkaan merupakan suatu kegiatan yang tidak dapat dipisahkan dari suatu penelitian. Teori-teori yang mendasari masalah dan bidang yang akan diteliti dapat ditemukan dengan melakukan studi pustaka. Selain itu seorang penelitian dapat memperoleh informasi tentang penelitian-penelitian sejenis atau yang ada kaitannya dengan penelitiannya. Dan penelitian-penelitian yang telah dilakukan sebelumnya..

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 siswa karena bersifat unik dan data lainnya.

Tabel 3.1 Sepuluh atribut pada data siswa baru SMK Bhakti Putra

No	Atribut	Keterangan
1	No	Nomor urut
2	NIS	Nomor Pokok Siswa
3	Nama Siswa	Nama Siswa
4	L/P	Jenis Kelamin L= Laki-Laki P= Perempuan
5	Jurusan	Program Studi 1. Rekayasa Perangkat Lunak 2. Bisnis Daring dan Pemasaran
6	Tanggal Lahir	Tanggal Lahir Siswa
7	Asal Sekolah	Asal Sekolah Siswa
8	Media	Media Promosi Sekolah 1. Brosur 2. Internet 3. Rekomendasi 4. Spanduk
9	Alamat	Alamat Siswa
10	Gaji Orang Tua	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

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.

Tabel 3.2 Data Kosong Pada Setiap Atribut Tahun 2017-2020

Atribut	**2017**	**2018**	**2019**	**2020**
NIS	0	0	0	0
Nama Siswa	0	0	0	0
L/P	0	0	0	0
Jurusan	1	2	2	2
Tanggal Lahir	2	3	2	0
Asal Sekolah	2	0	1	3
Program	0	2	1	2
Media	1	3	4	2
Alamat	2	0	0	1
Gaji Orang Tua	0	1	2	2

4. *Data Transformation*

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

Pengisian nilai kosong dapat menggunakan modus, *mean*, random. Modus merupakan pengisian data diisi dengan data yang paling banyak muncul seperti pada tabel dibawah ini

Tabel 3.3 Pengisian Data Kosong dengan Modus

No.	**Asal Sekolah**	**Media**
59.	MTS Alhidayah Ciaro	Rekomendasi
74.	MTS Alhidayah Ciaro	Brosur
137.	MTS Alhidayah Ciaro	Brosur
184.	MTS Alhidayah Ciaro	Spanduk
191.	MTS Alhidayah Ciaro	Rekomendasi

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 Index.

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 Index berupa teks dan tabel.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Selection* Data

Selection data digunakan untuk memilih data apa saja yang berpengaruh dan tidak berpengaruh dengan strategi promosi. Tabel 4.1 merupakan atribut-atribut pada data Siswa baru yang akan digunakan dan Tabel 4.2 merupakan atribut-aribut yang akan dihilangkan. Informasi didapat berdasarkan wawancara dengan bagian Kesiswaan Info PSB pada lampiran 1.

Tabel 4.1 Pemilihan Atribut yang Digunakan

Atribut Yang Digunakan		**Keterangan**
1.	Asal Sekolah	Untuk mengetahui sasaran promosi yang tepat
2.	Jurusan	Untuk mengetahui jurusan terbanyak peminatnya
3.	Media	Untuk media promosi yang digunakan
4.	Kecamatan	Untuk penempatan spanduk/ poster yang tepat
5.	Gaji Orang Tua	Untuk mengetahui gaji orang tua menengah ke bawah
6.	Jenis Kelamin	Untuk mengetahui Laki laki lebih memilih jurusan apa dan perempuan lebih dominan memilih jurusan apa

Tabel 4.2 Pemilihan Atribut yang Dihilangkan

Atribut Yang Dihilangkan		**Keterangan**
1.	No	Memiliki Nilai Unik
2.	NIS	Memiliki Nilai Unik
3.	Nama	Memiliki Nilai Unik
4.	Tempat Lahir	Dianggap sama dengan alamat

4.2 *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 4.3

Tabel 4.3 Pengisian Data Kosong dengan Modus

No.	Asal Sekolah	Media
13.	MTS Alhidayah Ciaro	Rekomendasi
102.	MTS Alhidayah Ciaro	Rekomendasi
113.	MTS Alhidayah Ciaro	Brosur
171.	MTS Alhidayah Ciaro	Brosur
172.	MTS Alhidayah Ciaro	Spanduk
183.	MTS Alhidayah Ciaro	Brosur
200.	MTS Alhidayah Ciaro	Brosur
250.	MTS Alhidayah Ciaro	Brosur
315.	MTS Alhidayah Ciaro	Rekomendasi

Tabel 4.3 pada *record* 200 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 2019 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 4.4.

Tabel 4.4 Pengisian Data Kosong dengan *Mean*

No.	Media	Asal Sekolah
4.	22	SMPN 3 Limbangan
23.	21	SMPN 3 Limbangan
34.	19	SMPN 3 Limbangan
97.	21	SMPN 3 Limbangan
115.	19	SMPN 3 Limbangan
119.	22	SMPN 3 Limbangan
141.	19	SMPN 3 Limbangan
173.	19	SMPN 3 Limbangan
174.	19	SMPN 3 Limbangan
195.	22	SMPN 3 Limbangan
247.	24	SMPN 3 Limbangan

Tabel 4.4 pada *record* 34 mempunyai data kosong pada atribut “”, kemudian akan digunakan penggunaan rata-rata yang diurutkan berdasarkan Asal Sekolah “SMPN 3 Limbangan” tahun 2017 sehingga nilai rata-ratanya 20,7 yang dibulatkan menjadi 21. *Random* merupakan pengisian data kosong yang diisi secara random seperti pada Tabel 4.5.

Tabel 4.5 Pengisian Data Kosong dengan *Random*

No.	Asal Sekolah	Media
126.	SMPN Ganjar Sabar	*Internet

Tabel 4.5 pada *record* 126 mempunyai data kosong pada aribut “Media”, kemudian akan digunakan penggunaan random karena Siswa yang asal sekolahnya SMPN Ganjar Sabar hanya ada satu, sehingga data kosong diisi dengan data media “Internet”. Contoh pengisian data tidak konsisten seperti pada Tabel 4.6.

Tabel 4.6 Perbaikan Data Tidak Konsisten

No.	Asal Sekolah
109.	SMPN 4 Limbangan
136.	SMPN 4 Limbangan
166.	SMPN 4 Limbangan
255.	SMPN 4 Limbangan

Tabel 4.6 yang telah diurutkan berdasarkan asal sekolah sesuai abjad A-Z, kolom asal sekolah pada *record* no 136 diisi dengan “SMPN 4 Limbangan”, sedangkan pada data asal sekolah lain diisi dengan “SMP 1 Nagreg” sehingga asal sekolah pada *record* no 136 harus diubah agar datanya menjadi konsisten. Sehingga data akan diperbaiki menjadi “SMPN 4 Limbangan”. Contoh perbaikan data dalam salah pengetikan data seperti Tabel 4.7.

Tabel 4.7 Perbaikan Data dalam Pengetikan Data

No.	Asal Sekolah
214.	SMPN 4 Limbangan
306.	SMPN 4 Limbangan
327.	SMPN 4 Limbangan
392.	SMPN 4 Limbangan
402.	SMPN 4 Limbangan

Gambar 4.7 telah diurutkan berdasarkan asal sekolah sesuai abjad A-Z, kolom asal sekolah pada *record* no 214 kolom asal sekolah diisi dengan “SMPN 1 Nagreg”, sedangkan pada data asal sekolah lain diisi dengan “SMPN 1 Nagreg” sehingga asal sekolah pada *record* no 306 harus diubah agar datanya menjadi konsisten. Sehingga data diperbaiki menjadi “SMPN 1 Nagreg”. Apabila data kosong dalam satu *record* masih banyak, maka dapat dilakukan penghapusan satu *record* seperti pada Tabel 4.8. yang mempunyai satu data kosong.

Tabel 4.8 Menghilangkan Data Yang Mempunyai Banyak Data Kosong

No	Jurusan	Asal Sekolah	Media	Alamat	Gaji Ortu
57.					

Setelah tidak ada data yang kosong, maka data Siswa baru diintegrasikan dengan cara menggabungkan tabel Siswa dari tahun 2017 sampai dengan tahun 2020 menjadi satu tabel. Jumlah total hasil akhir *record* data Siswa baru setelah dilakukan proses pembersihan data adalah 520 *record*. Hasil integrasi tabel seperti pada Tabel 4.9.

Tabel 4.9 Hasil Integrasi Tabel Tahun 2017 sampai 2020

No	Jurusan	Asal Sekolah	Media	Kecamatan	Gaji Orang Tua	L/P
1.	RPL	MTS Fatih Bandung	Spanduk	Nagreg	1.450.000	Perempuan
2.	RPL	MTS Pulosari	Brosur	Limbangan	1.450.000	Laki Laki
3.	RPL	SMPN 3 Limbangan	Brosur	Nagreg	2.450.000	Laki Laki
4.	RPL	SMPN 3 Limbangan	Internet	Nagreg	2.450.000	Laki Laki
5.	RPL	MTS Al Furqon	Brosur	Nagreg	3.450.000	Laki Laki
6.	RPL	MTS Fatih Bandung	Rekomendasi	Nagreg	1.450.000	Perempuan
7.	RPL	MTS Alhidayah	Brosur	Nagreg	2.450.000	Laki Laki
8.	RPL	MTS Alhidayah	Brosur	Nagreg	1.450.000	Laki Laki
9.	RPL	MTS Alhidayah	Brosur	Nagreg	1.450.000	Perempuan
10.	RPL	MTS Alhidayah	Brosur	Nagreg	1.450.000	Laki Laki
...	...	...	...	...	...	...
520.	RPL	MTS Alhidayah	Brosur	RPL	1.450.000	Laki Laki

Data Siswa hasil integrasi pada Lampiran 3.

4.3 Transformasi Data

Atribut alamat akan di ambil data Kecamatan sesuai kebutuhan Sekolah. Tabel 4.10 menampilkan hasil atribut alamat yang sudah dipisah berdasarkan Kecamatan.

Tabel 4.10 Transformasi Alamat menjadi Kecamatan

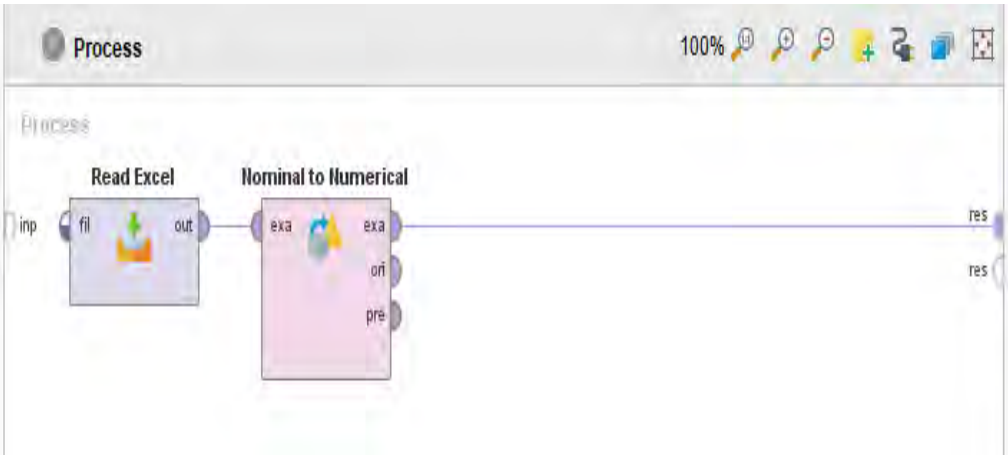
No.	Alamat	Hasil Transformasi
1.	Ciaro Kab. Bandung	Nagreg
2.	Cijolang, Garut	Limbangan
3.	Jalan Cagak. Bandung	Nagreg
4.	Andir Kab.Bandung	Nagreg
5.	Lebak Candra Desa Simpen Kab. Garut	Limbangan
6.	Tegalame Kab.Bandung	Nagreg
7.	Ciaro Kab. Garut	Nagreg
...		...
520.	Ciaro Kab. Bandung	Nagreg

Tabel 4.11 merupakan Atribut “Gaji Orang Tua” yang ditransformasikan menggunakan rata-rata.

Tabel 4.11 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

Tahapan selanjutnya adalah dengan mentransformasikan data dari data yang berbentuk nominal menjadi numerik. 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 menjadi 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 1 sampai tidak terhingga. Inisialisasi dilakukan sesuai dengan urutan muculnya data, contohnya pada atribut Jurusan pada Tabel 4.12 Transformasi Data Atribut Jurusan

No	Jurusan	Hasil Transformasi
1.	RPL	1
2.	RPL	1
3.	RPL	1
4.	RPL	1
5.	RPL	1
6.	RPL	1
7.	BDP	2
8.	BDP	2
9.	BDP	2
10.	BDP	2
11.	BDP	2
12.	BDP	2

Berdasarkan Gambar 4.1, Program Studi yang keluar pertama adalah “RPL” sehingga data di inisialisasi dengan angka 1 dan “BDP” dengan angka 2. Sehingga data siswa baru tahun 2017 sampai dengan 2020 setelah di *transformasi* seperti Tabel 4.9.

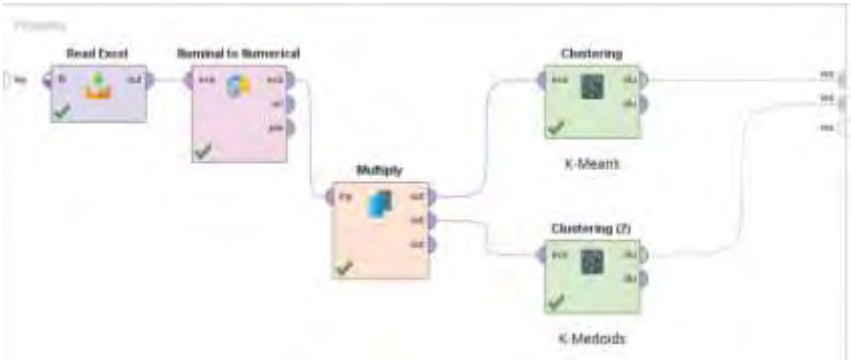
Tabel 4.13 Data Hasil Transformasi Pada Data Siswa Baru

No	Jurusan	Asal Sekolah	Media	Kecamatan	Jenis Kelamin	Gaji Orang Tua (Rp.)
1.	1	3	2	1	2	1.450.000
2.	1	15	1	2	1	1.450.000
3.	1	2	1	1	1	2.450.000
4.	1	2	3	1	1	2.450.000
5.	1	4	1	2	1	3.450.000
6.	1	3	4	1	1	1.450.000
7.	1	1	1	1	2	2.450.000
8.	1	1	1	1	1	1.450.000
9.	1	1	1	1	2	1.450.000
10.	1	1	1	1	1	1.450.000
...	...	...	...	...	...	...
520.	1	1	1	1	1	1.450.000

Sumber : Pengolahan data oleh Penulis

4.4 *Implementasi Algoritma Clustering*

Untuk menggali informasi yang diambil dari Tabel 4.13, berupa data calon Siswa baru Tahun 2017 sampai dengan 2020 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 *cluster*ing pertama dan *K-Medoids* pada *clustering* kedua. Parameter *clustering K-Means* yang digunakan dalam penelitian ini seperti pada Tabel 4.14.

Tabel 4.14 Parameter *clustering K-Means*

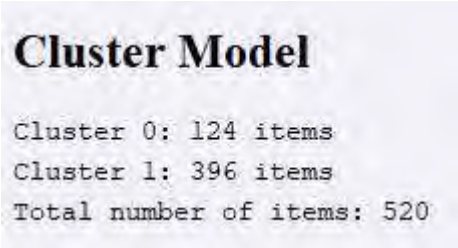
Parameter	Value
K	2 sampai dengan 7
Max runs	8 (*default*)
Measure type	*Numerical measurement*
Numerical measure	*EuclideanDistance*
Max optimization step	100 (*default*)

Parameter *clustering K-Medoids* yang digunakan dalam penelitian ini seperti pada Tabel 4.15.

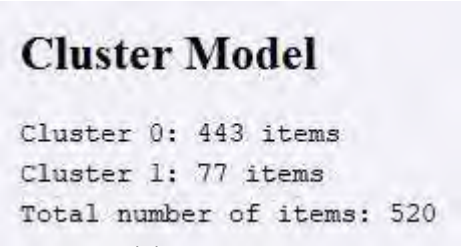
Tabel 4.15 Parameter *Clustering K-Medoids*

Parameter	Value
K	2 sampai dengan 7
Max runs	8 (*default*)
Max Optimization Step	100
Measure type	*Numerical Measurement*
Numerical measure	*Euclide anDistance*
Max optimization step	100 (*default*)

Perbanding *Cluster* Model dari K=2 sampai dengan K=7 seperti pada Gambar 4.3 sampai dengan 4.8.



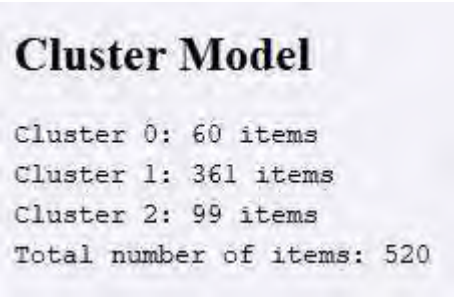
(a)



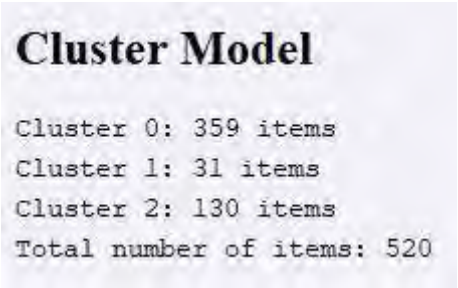
(b)

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 124 *record* dan *cluster* 1 berjumlah 396 *record*. Sedangkan algoritma *K-Medoids* K=2 pada *cluster* 0 sebesar 443 *record* dan *cluster* 1 berjumlah 77 *record*.



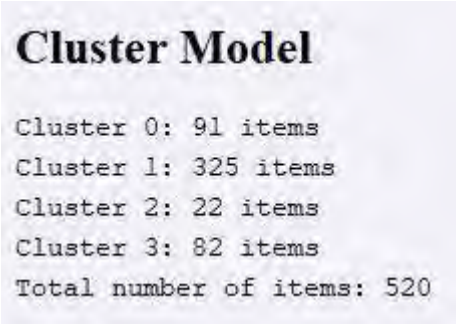
(a)



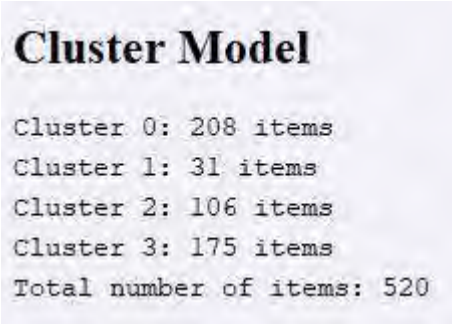
(b)

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 60 *record*, *cluster* 1 berjumlah 361 *record* dan *cluster* 2 berjumlah 99 *record*. Sedangkan algoritma *K-Medoids* K=3 pada *cluster* 0 sebesar 359 *record*, *cluster* 1 berjumlah 31 *record* dan *cluster* 2 berjumlah 130 *record*.



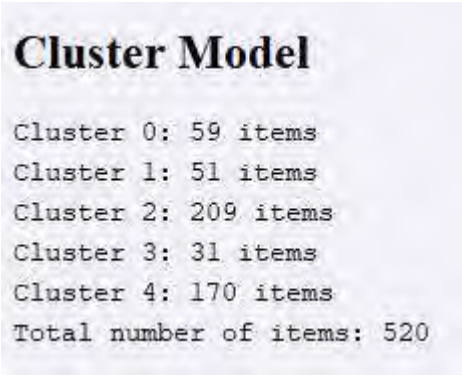
(a)



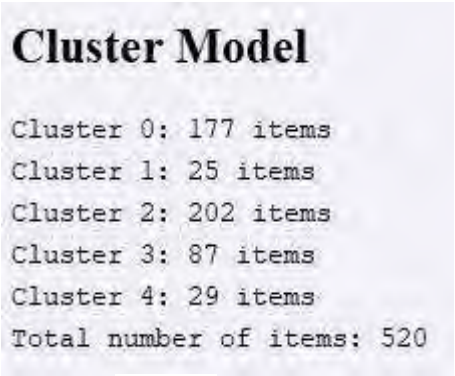
(b)

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 91 *record*, *cluster* 1 berjumlah 325 *record*, *cluster* 2 berjumlah 22 *record* dan *cluster* 3 berjumlah 82 *record*. Sedangkan algoritma *K-Medoids* K=4 pada *cluster* 0 sebesar 208 *record*, *cluster* 1 berjumlah 31 *record*, *cluster* 2 berjumlah 106 *record* dan *cluster* 3 berjumlah 175 *record*.



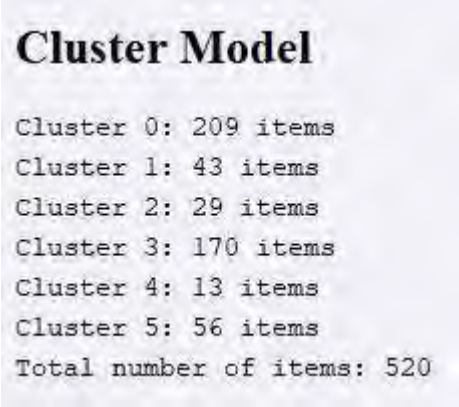
(a)



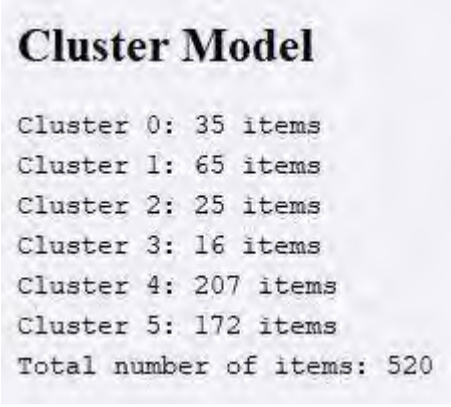
(b)

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 59 *record*, *cluster* 1 berjumlah 51 *record*, *cluster* 2 berjumlah 209 *record*, *cluster* 3 berjumlah 31 *record* dan *cluster* 4 berjumlah 170 *record*. Sedangkan algoritma *K-Medoids* K=5 pada *cluster* 0 sebesar 177 *record*, *cluster* 1 berjumlah 25 *record*, *cluster* 2 berjumlah 202 *record, cluster* 3 berjumlah 87 *record* dan *cluster* 4 berjumlah 29 *record*.



(a)

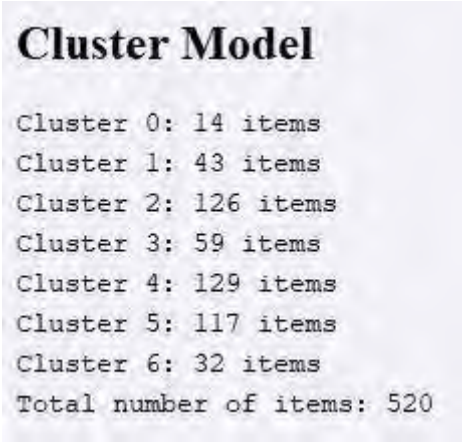


(b)

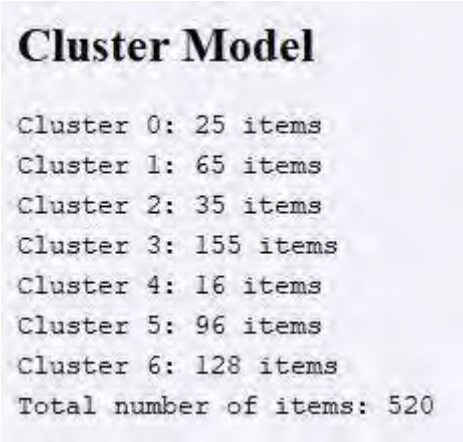
Gambar 4.7
Cluster model K = 6 pada *K-Means* (a) dan *K-Medoids* (b)

Berdasarkan Gambar 4.7, jumlah data pengelompokkan data K=6 pada Algoritma *K-Means* di *cluster* 0 berjumlah 209 *record*, *cluster* 1 berjumlah 43 *record*, *cluster* 2 berjumlah 29 *record*, *cluster* 3 berjumlah 170 *record, cluster* 4 berjumlah 13 *record,* dan *cluster* 5 berjumlah 56 *record*. Sedangkan algoritma *K-Medoids* K=6 pada *cluster* 0 sebesar

35 *record*, *cluster* 1 berjumlah 65 *record*, *cluster* 2 berjumlah 25 *record, cluster* 3 berjumlah 16 *record, cluster* 4 berjumlah 207 *record* dan *cluster* 5 berjumlah 172 *record*.



(a)



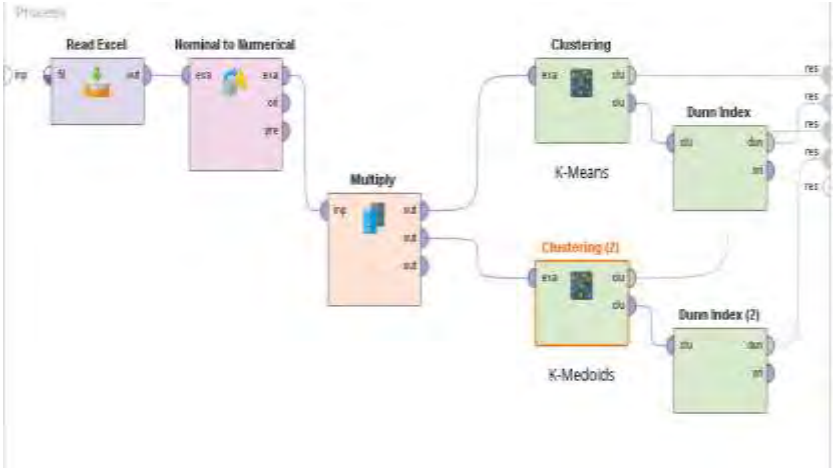
(b)

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 14 *record*, *cluster* 1 berjumlah 43 *record*, *cluster* 2 berjumlah 126 *record*, *cluster* 3 berjumlah 59 *record, cluster* 4 berjumlah 129 *record, cluster* 5 berjumlah 117 *record* dan *cluster* 6 berjumlah 32 *record*. Sedangkan algoritma *K-Medoids* K=7 pada *cluster* 0 sebesar 25 *record* , *cluster* 1 berjumlah 65 *record*, *cluster* 2 berjumlah 35 *record, cluster* 3 berjumlah 155 *record, cluster* 4 berjumlah 16 *record, cluster* 5 berjumlah 96 *record* dan *cluster* 6 berjumlah 128 *record*.

4.5 *Pattern Evaluation*

Setelah mendapatkan pola-pola unik pada percobaan K=2 sampai dengan K=7, maka untuk mengetahui mana K terbaik adalah dengan menggunakan operator validasi *dunn Index* seperti Gambar 4.9



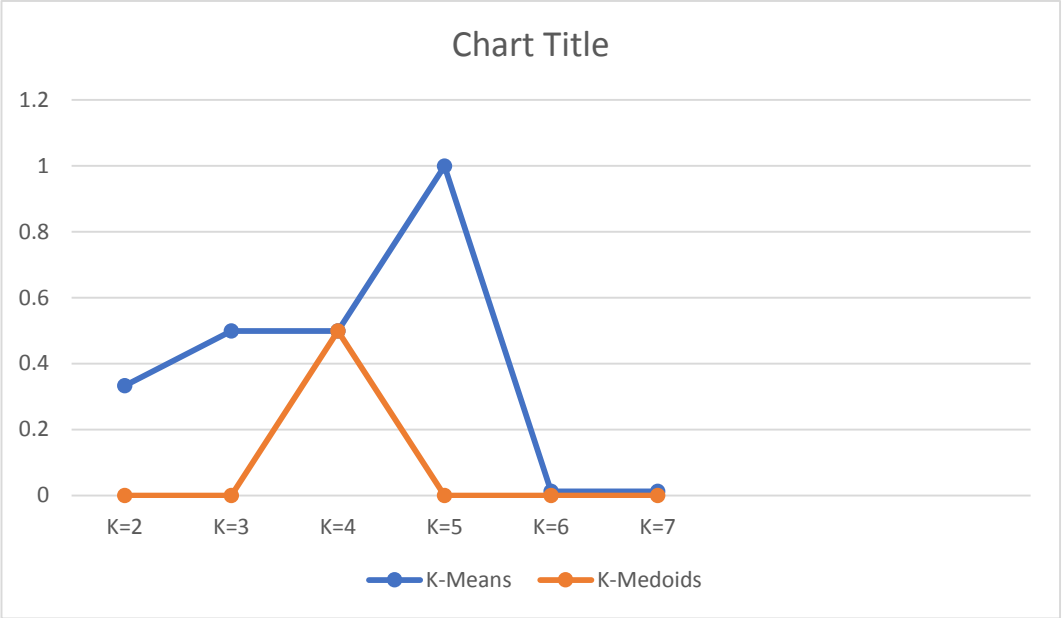
Gambar 4.9
Penambahan operator validasi dunn *Index*

Setelah melakukan percobaan terhadap parameter tersebut pada Tabel 4.16 diperoleh Nilai *Dunn Index* dan lama proses *running.*

Tabel 4.16 *Cluster K-Means* dan *K-Medoids* terhadap Nilai *Dunn Index*

Cluster	***K-Means***	**Waktu**	***K-Medoids***	**Waktu**
2	0.333	1s	∞	49s
3	0.499	1s	∞	50s
4	0.499	1s	0.488	55s
5	0.988	1s	∞	59s
6	∞	1s	∞	1m 05s
7	0,013	1s	∞	1m 08s

Gambar 4.10 merupakan grafik berdasarkan Tabel 4.16 sehingga nilai *dunn Index* mudah untuk dipahami.



Gambar 4.10
Grafik *Cluster K-Means* dan *K-Medoids* terhadap Nilai Dunn *Index*

Berdasarkan (Chowdary N. S., 2014) yang menyatakan bahwa semakin besar Indexnya, semakin baik hasil pengelompokannya, maka dilihat dari Gambar 4.10, 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

k=5 seperti Gambar 4.11.

Cluster Model

```
Cluster 0: 59 items
Cluster 1: 51 items
Cluster 2: 209 items
Cluster 3: 31 items
Cluster 4: 170 items
Total number of items: 520
```

Gambar 4.11

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

Gambar 4.11 menjelaskan jumlah data calon siswa baru pada setiap *cluster*nya yang terbagi menjadi 5 *cluster*, dimana *cluster* 0 dengan jumlah 59 data, *cluster* 1 dengan 51 data, *cluster* 2 dengan 209 data, *cluster* 3 dengan 31 data dan *cluster* 4 dengan 170 data. Tabel 4.17 merupakan hasil perhitungan antara jarak *cluster* dengan *centroid* menggunakan *Rapidminer*.

Tabel 4.17 Hasil Perhitungan Jarak *Cluster* dengan *Centroid*

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
L/P	1.340	1.442	1.379	1.424	1.231
Jurusan	1.340	1.442	1.552	1.459	1.308
Asal Sekolah	1.818	23,860	11.448	5.759	28.923
Media	1.651	1.512	1.724	1.812	1.462
Alamat	1.292	4.558	13.069	2.806	14.538
Gaji Orang Tua	1.435	1.605	1.828	1.447	1.154

Sumber : Rapidminer

4.6 *Interpretation Knowledge*

Penjabaran dari *Cluster* Model pada Gambar 4.11. dengan nilai K=5 pada Algoritma *K-Means* seperti Tabel 4.17 yang berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data siswa pada setiap atribut. seperti pada Tabel 4.18 sampai Tabel 4.22

Tabel 4.18 Hasil Analisis *Cluster* 0

***Cluster* 0**							
Program Studi						**Media Promosi**	
Rekayasa Perangkat Lunak	36					Brosur	37
Bisnis Daring Dan Pemasaran	23					Spanduk	12
Jenis Kelamin						Internet	9
Laki Laki	37					Rekomendasi	1
Perempuan	22						
Asal Sekolah							
SMPN 1 Limbangan	10		SMPN 1 Nagreg	2		MTS Alhidayah Ciaro	1
SMPN 2 Limbangan	8		SMPN 3 Kadungora	1		MTS Fatih Bandung	1
SMPN 3 Limbangan	1		SMP PGRI Ganjar Sabar	1		MTS Al Furqon	5
SMPN 4 Limbangan	1		SMPN 1 Cicalengka	1		MTS Mathalul Anwar	2
SMPN Tribakti Nagreg	5		SMPN 2 Cicalengka	1		MTS Pulosari	8
SMP YP 17 Nagreg	0		SMPN Bojong	1		MTS Alihsan BL. Limbangan	0
SMPN 1 Kadungora	0		SMPN PGRI Limbangan	!		MTS Mahmud BL. Limbangan	1
SMP Al MAMUN Limbangan	3		SMP IT Miftahul Khoer	8		MTS Muhamadiyah 9	!
SMP Al Fatah Limbangan	1		SMP Wiraguna	8		MTS Muhamadiyah Cisaat	3
SMP As Salwa BL. Limbangan	!		MTS Miftahul Janah	1		MTS As Syukur	2
SMP Maarif Cicalengka	!		SMP PGRI Cicalengka	3		MTS Al Baroroh	1
SMP F.K Bina Muda	1		SMP Yadika Cicalengka	3		MTS Sunan Cipancar BL Limbangan	!
Kecamatan							
Nagreg	8		Rancaekek	1		Pacet	2
Limbangan	38		Cibugel	1		Majalaya	2
Selaawi	1		Cibiuk	4		Haur Pugur	1
Kadungora	1		Leuwigoong	1		Solokan Jeruk	1
Cicalengka	6		Cibatu	0		Malangbong	1
Leles	1		Cikancung	1		Tanjungsari	1
Cangkuang	1		Paseh	0		Jatinangor	1
Rata-Rata Gaji Orang Tua							
Rp. 1.200.000							

Berdasarkan Tabel 4.18 *Cluster* 0 memiliki karakteristik calon siswa yang mempunyai rata-rata gaji orang tua sebesar Rp. 1.200.000 dengan Rekayasa perangkat lunak menjadi program studi yang paling banyak diminati dengan persentase 61% dan Bisnis daring dan pemasaran menjadi Jurusan yang peminatnya lebih sedikit dengan persentase 39%. *Cluster* 0 kebanyakan berasal dari SMPN 1 Limbangan dengan persentase 16,9%. Media brosur dan spanduk menjadi media promosi paling besar dengan perbedaan persentase 42,4% dan Kecamatan BL Limbangan menjadi kecamatan paling banyak dengan persentase 64,4% dan jenis kelamin Pria lebih banyak dibandingkan perempuan dengan presentasi 62,7%.

Tabel 4.19 Hasil Analisis *Cluster* 1

***Cluster* 1**							
Program Studi						**Media Promosi**	
Rekayasa Perangkat Lunak	29					Brosur	13
Bisnis Daring Dan Pemasaran	22					Spanduk	8
Jenis Kelamin						Internet	4
Laki Laki	32					Rekomendasi	32
Perempuan	19						
Asal Sekolah							
SMPN 1 Limbangan	1		SMPN 1 Nagreg	1		MTS Alhidayah Ciaro	1
SMPN 2 Limbangan	1		SMPN 3 Kadungora	1		MTS Fatih Bandung	1
SMPN 3 Limbangan	1		SMP PGRI Ganjar Sabar	1		MTS Al Furqon	1
SMPN 4 Limbangan	1		SMPN 1 Cicalengka	2		MTS Mathalul Anwar	1
SMPN Tribakti Nagreg	1		SMPN 2 Cicalengka	2		MTS Pulosari	0
SMP YP 17 Nagreg	3		SMPN Bojong	3		MTS Alihsan BL. Limbangan	2
SMPN 1 Kadungora	3		SMPN PGRI Limbangan	1		MTS Mahmud BL. Limbangan	3
SMP Al MAMUN Limbangan	0		SMP IT Miftahul Khoer	1		MTS Muhamadiyah 9	1
SMP Al Fatah Limbangan	5		SMP Wiraguna	1		MTS Muhamadiyah Cisaat	1
SMP As Salwa BL. Limbangan	6		MTS Miftahul Janah	6		MTS As Syukur	5
SMP Maarif Cicalengka	1		SMP PGRI Cicalengka	1		MTS Al Baroroh	4

***Cluster* 1**							
SMP F.K Bina Muda	3		SMP Yadika Cicalengka	1		MTS Sunan Cipancar BL Limbangan	0
Kecamatan							
Nagreg	6		Rancaekek	1		Pacet	1
Limbangan	9		Cibugel	1		Majalaya	1
Selaawi	7		Cibiuk	5		Haur Pugur	2
Kadungora	4		Leuwigoong	2		Solokan Jeruk	1
Cicalengka	2		Cibatu	1		Malangbong	2
Leles	1		Cikancung	8		Tanjungsari	1
Cangkuang	1		Paseh	3		Jatinangor	1
Rata-Rata Gaji Orang Tua							
Rp. 1.500.000							

Berdasarkan Tabel 4.19 *Cluster* 1 memiliki karakteristik calon siswa yang mempunyai rata-rata gaji orang tua sebesar Rp. 1.500.000 dengan bisnis daring dan pemasaran menjadi jurusan yang paling banyak diminati dengan persentase 56,9% dan rekayasa perangkat lunak jurusan yang peminatnya lebih sedikit dengan persentase 43,1%. *Cluster* 1 kebanyakan berasal dari SMP As Salwa Limbangan dan MTs Miftahul Janah dengan persentase 11,9%. Media brosur dan spanduk menjadi media promosi paling besar dengan perbedaan persentase 22,4% dan Kecamatan Cikancung menjadi kecamatan paling banyak dengan persentase 15,7% dan jenis kelamin Pria lebih banyak dibandingkan perempuan dengan presentasi 62,7%.

Tabel 4.20 Hasil Analisis *Cluster* 2

***Cluster* 2**							
Program Studi						**Media Promosi**	
Rekayasa Perangkat Lunak	138					Brosur	137
Bisnis Daring Dan Pemasaran	71					Spanduk	30
Jenis Kelamin						Internet	20
Laki Laki	138					Rekomendasi	22
Perempuan	71						
Asal Sekolah							
SMPN 1 Limbangan	1		SMPN 1 Nagreg	1		MTS Alhidayah Ciaro	100
SMPN 2 Limbangan	!		SMPN 3 Kadungora	1		MTS Fatih Bandung	52

***Cluster* 2**							
SMPN 3 Limbangan	52		SMP PGRI Ganjar Sabar	1		MTS Al Furqon	5
SMPN 4 Limbangan	1		SMPN 1 Cicalengka	1		MTS Mathalul Anwar	0
SMPN Tribakti Nagreg	1		SMPN 2 Cicalengka	1		MTS Pulosari	0
SMP YP 17 Nagreg	1		SMPN Bojong	1		MTS Alihsan BL. Limbangan	0
SMPN 1 Kadungora	1		SMPN PGRI Limbangan	0		MTS Mahmud BL. Limbangan	0
SMP Al MAMUN Limbangan	1		SMP IT Miftahul Khoer	0		MTS Muhamadiyah 9	0
SMP Al Fatah Limbangan	1		SMP Wiraguna	0		MTS Muhamadiyah Cisaat	0
SMP As Salwa BL. Limbangan	1		MTS Miftahul Janah	0		MTS As Syukur	0
SMP Maarif Cicalengka	1		SMP PGRI Cicalengka	0		MTS Al Baroroh	0
SMP F.K Bina Muda	1		SMP Yadika Cicalengka	0		MTS Sunan Cipancar BL Limbangan	0
Kecamatan							
Nagreg	137		Rancaekek	1		Pacet	0
Limbangan	19		Cibugel	1		Majalaya	1
Selaawi	1		Cibiuk	2		Haur Pugur	0
Kadungora	1		Leuwigoong	3		Solokan Jeruk	0
Cicalengka	12		Cibatu	1		Malangbong	0
Leles	1		Cikancung	1		Tanjungsari	1
Cangkuang	0		Paseh	1		Jatinangor	1
Rata-Rata Gaji Orang Tua							
Rp. 3.000.000							

Berdasarkan Tabel 4.20 *Cluster* 2 memiliki karakteristik calon siswa yang mempunyai rata-rata gaji orang tua sebesar Rp. 3.000.000 dengan Rekayasa perangkat lunak menjadi program studi yang paling banyak diminati dengan persentase 66% dan Bisnis daring dan pemasaran menjadi Jurusan yang peminatnya lebih sedikit dengan persentase 34%. *Cluster* 2 kebanyakan berasal dari MTS Alhidayah Ciaro dengan persentase 47,8%. Media brosur dan spanduk menjadi media promosi paling besar dengan perbedaan persentase 51,2% dan Kecamatan Nagreg menjadi kecamatan paling banyak dengan persentase 82,8% dan jenis kelamin Pria lebih banyak dibandingkan perempuan dengan presentasi 66%.

Tabel 4.21 Hasil Analisis *Cluster* 3

***Cluster* 3**							
Program Studi						**Media Promosi**	
Rekayasa Perangkat Lunak	15					Brosur	12
Bisnis Daring Dan Pemasaran	16					Spanduk	24
Jenis Kelamin						Internet	4
Laki Laki	19					Rekomendasi	4
Perempuan	12						
Asal Sekolah							
SMPN 1 Limbangan	1		SMPN 1 Nagreg	1		MTS Alhidayah Ciaro	1
SMPN 2 Limbangan	1		SMPN 3 Kadungora	3		MTS Fatih Bandung	1
SMPN 3 Limbangan	1		SMP PGRI Ganjar Sabar	1		MTS Al Furqon	0
SMPN 4 Limbangan	1		SMPN 1 Cicalengka	1		MTS Mathalul Anwar	0
SMPN Tribakti Nagreg	1		SMPN 2 Cicalengka	1		MTS Pulosari	0
SMP YP 17 Nagreg	1		SMPN Bojong	0		MTS Alihsan BL. Limbangan	0
SMPN 1 Kadungora	2		SMPN PGRI Limbangan	0		MTS Mahmud BL. Limbangan	1
SMP Al MAMUN Limbangan	4		SMP IT Miftahul Khoer	0		MTS Muhamadiyah 9	0
SMP Al Fatah Limbangan	1		SMP Wiraguna	0		MTS Muhamadiyah Cisaat	0
SMP As Salwa BL. Limbangan	1		MTS Miftahul Janah	0		MTS As Syukur	0
SMP Maarif Cicalengka	6		SMP PGRI Cicalengka	8		MTS Al Baroroh	0
SMP F.K Bina Muda	1		SMP Yadika Cicalengka	7		MTS Sunan Cipancar BL Limbangan	0
Kecamatan							
Nagreg	1		Rancaekek	1		Pacet	1
Limbangan	1		Cibugel	4		Majalaya	3
Selaawi	1		Cibiuk	1		Haur Pugur	1
Kadungora	!		Leuwigoong	1		Solokan Jeruk	5
Cicalengka	1		Cibatu	3		Malangbong	0
Leles	1		Cikancung	1		Tanjungsari	7
Cangkuang	2		Paseh	1		Jatinangor	3
Rata-Rata Gaji Orang Tua							
Rp. 2.000.000							

Berdasarkan Tabel 4.21 *Cluster* 3 memiliki karakteristik calon siswa yang mempunyai rata-rata gaji orang tua sebesar Rp. 2.000.000 dengan Bisnis Daring dan Pemasaran menjadi program studi yang paling banyak diminati dengan persentase 51,6% dan Rekayasa perangkat lunak menjadi Jurusan yang peminatnya lebih sedikit dengan persentase 48,4%. *Cluster* 3 kebanyakan berasal dari SMP PGRI Cicalengka dengan persentase 25,8%. Media spanduk menjadi media promosi paling besar dengan persentase 61,3% dan Kecamatan Tanjung Sari menjadi kecamatan paling banyak dengan persentase 22,6% dan jenis kelamin Pria lebih banyak dibandingkan perempuan dengan presentasi 38,7%.

Tabel 4.22 Hasil Analisis *Cluster* 4

***Cluster* 4**							
Program Studi						**Media Promosi**	
Rekayasa Perangkat Lunak	92					Brosur	95
Bisnis Daring Dan Pemasaran	78					Spanduk	27
Jenis Kelamin						Internet	33
Laki Laki	98					Rekomendasi	15
Perempuan	72						
Asal Sekolah							
SMPN 1 Limbangan	0		SMPN 1 Nagreg	0		MTS Alhidayah Ciaro	0
SMPN 2 Limbangan	0		SMPN 3 Kadungora	42		MTS Fatih Bandung	0
SMPN 3 Limbangan	3		SMP PGRI Ganjar Sabar	0		MTS Al Furqon	41
SMPN 4 Limbangan	40		SMPN 1 Cicalengka	0		MTS Mathalul Anwar	0
SMPN Tribakti Nagreg	0		SMPN 2 Cicalengka	0		MTS Pulosari	0
SMP YP 17 Nagreg	0		SMPN Bojong	1		MTS Alihsan BL. Limbangan	0
SMPN 1 Kadungora	0		SMPN PGRI Limbangan	12		MTS Mahmud BL. Limbangan	0
SMP Al MAMUN Limbangan	2		SMP IT Miftahul Khoer	1		MTS Muhamadiyah 9	15
SMP Al Fatah Limbangan	2		SMP Wiraguna	1		MTS Muhamadiyah Cisaat	0
SMP As Salwa BL. Limbangan	2		MTS Miftahul Janah	1		MTS As Syukur	0
SMP Maarif Cicalengka	5		SMP PGRI Cicalengka	1		MTS Al Baroroh	0
SMP F.K Bina Muda	2		SMP Yadika Cicalengka	1		MTS Sunan Cipancar BL Limbangan	12
Kecamatan							

***Cluster* 4**							
Nagreg	34		Rancaekek	0		Pacet	0
Limbangan	81		Cibugel	0		Majalaya	0
Selaawi	3		Cibiuk	17		Haur Pugur	0
Kadungora	9		Leuwigoong	9		Solokan Jeruk	0
Cicalengka	10		Cibatu	0		Malangbong	0
Leles	7		Cikancung	0		Tanjungsari	0
Cangkuang	0		Paseh	0		Jatinangor	0
Rata-Rata Gaji Orang Tua							
Rp. 4.000.000							

Berdasarkan Tabel 4.22 *Cluster* 4 memiliki karakteristik calon siswa yang mempunyai rata-rata gaji orang tua sebesar Rp. 4.000.000 dengan Rekayasa perangkat lunak menjadi program studi yang paling banyak diminati dengan persentase 54,1% dan Bisnis daring dan pemasaran menjadi Jurusan yang peminatnya lebih sedikit dengan persentase 45,9%. *Cluster* 4 kebanyakan berasal dari SMPN 3 Kadungora dengan persentase 24,7%. Media brosur dan Internet menjadi media promosi paling besar dengan perbedaan persentase 36,5% dan Kecamatan BL Limbangan menjadi kecamatan paling banyak dengan persentase 47,6% dan jenis kelamin Pria lebih banyak dibandingkan perempuan dengan presentasi 57,6%.

4.23 merupakan karakerisik dari setiap *cluster*.

Tabel 4.23 Karakteristik setiap *cluster*

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Rata-rata gaji orang tua Rp. 1.000.000	Rata-rata gaji orang tua Rp. 1.500.000	Rata-rata gaji orang tua Rp. 3.000.000	Rata-rata gaji orang tua Rp. 2.000.000	Rata-rata gaji orang tua Rp. 4.000.000
RPL banyak diminati dengan 61,1 % Sedangkan BDP 39%	BDP banyak diminati dengan 56,9 % dan RPL 43,1%	RPL banyak diminati dengan 66 % dan RPL 34%	BDP banyak diminati dengan 51,6 % dan BDP 48,4%	RPL banyak diminati dengan 54,1 % dan BDP 45,9%

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Brosur dan Spanduk sangat berpegaruh dengan perbedaan persentase 42,4%	Rekomendasi sangat berpegaruh dengan persentase 47,8%	Brosur sangat berpegaruh dengan persentase 65,6%	Spanduk sangat berpegaruh dengan perbedaan persentase 61%	Brosur sangat berpegaruh dengan persentase 55,9%
SMPN 1 Limbangan 16,9% memberikan lulusan sebanyak 10 calon siswa	SMPN As Salwa Limbangan dan MTS Miftahul Jannah 11,2% memberikan lulusan sebanyak 11 calon siswa	MTS Alhidayah Ciaro 47,8% memberikan lulusan sebanyak 100 calon siswa dengan persentasi 47.8%	SMP PGRI cicalengka dan SMP yadika Cicalengka 11,2% memberikan lulusan sebanyak 8 calon siswa	SMPN 3 Kadungora 5,2% memberikan lulusan sebanyak 42 calon siswa dengan persentasi 24.7%
Kecamatan Limbangan paling banyak dengan persentase 64,4%	Kecamatan Cikancung paling banyak dengan persentase 15,7%	Kecamatan Nagregpaling banyak dengan persentase 82%	Kecamatan Tanjungsari paling banyak dengan persentase 22,6%	Kecamatan BL Limbangan paling banyak dengan persentase 47,6%
Jenis Kelamin Laki Laki Lebih banyak jika dibandingkan dengan perempuan	Jenis Kelamin Laki Laki Lebih banyak jika dibandingkan dengan perempuan	Jenis Kelamin Laki Laki Lebih banyak jika dibandingkan dengan perempuan	Jenis Kelamin Laki Laki Lebih banyak jika dibandingkan dengan perempuan	Jenis Kelamin Laki Laki Lebih banyak jika dibandingkan dengan perempuan

4.7 Strategi Promosi

Berdasarkan hasil pengelompokkan yang terbagi menjadi 5 *cluster*, dapat dilakukan strategi promosi berdasarkan *promotion mix* pada Tabel 4.24 menggunakan atribut-atribut yang saling berkaitan dan hasil wawancara dengan bagian Panitia PSB.

Tabel 4.24 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		√	√		√

No	Strategi Promosi	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	Persentasi Sekolah	√	√		√	
3	**Promosi Penjualan**					
	Pemberian Beasiswa	√	√		√	
4	**Hubungan Masyarakat**					
	Membuka Stand di Kecamatan	√	√	√	√	
	Mengadakan Perlombaan	√	√	√	√	√
	Melakukan Kerjasama Sekolah	√	√	√	√	√
5	**Pemasaran Langsung**					
	Internet	√	√	√	√	√

Berdasarkan hasil dari analisis bahwa ditemukan beberapa kesimpulan dari cluster 0 yaitu dari segi periklanan atau media promosi brosur dan spanduk harus diterapkan karena memiliki persentase yang cukup tinggi yakni 62,7% dan 20,3% sehingga perbedaan persentasenya 42,4% sedangkan untuk rekomendasi tidak diutamakan di *cluster* ini karena persentase yang sedikit, pemberian beasiswa biasanya dilakukan di cluster ini dikarenakan pendapatan orang tua yang paling sedikit dengan rata rata 1.200.000 dan membuka *stand* di beberapa kecamatan yang dominan. Untuk Hubungan masyarakat dan internet promosi juga dapat dijadikan strategi karena bersifat global.

Dari *cluster* 1 didapat data analisis yaitu Berdasarkan hasil dari analisis bahwa ditemukan beberapa kesimpulan dari *cluster* 1 yaitu dari segi periklanan atau media rekomendasi dan brosur harus diterapkan karena memiliki persentase yang cukup tinggi yakni 47,8% untuk rekomendasi dan 34,3% sehingga perbedaan persentasenya 22,4% sedangkan untuk rekomendasi tidak diutamakan di *cluster* ini karena persentase yang sedikit, pemberian beasiswa bias dilakukan di *cluster* ini dikarenakan pendapatan orang tua yang paling sedikit dengan rata rata 1.200.000 dan membuka *stand* di beberapa kecamatan yang dominan. Untuk Hubungan masyarakat dan internet promosi juga dapat dijadikan strategi karena bersifat global.

Berdasarkan hasil dari analisis bahwa ditemukan beberapa kesimpulan dari cluster 0 yaitu dari segi periklanan atau media promosi brosur dan spanduk harus diterapkan karena memiliki presentasi yang cukup tinggi yakni 23% dan 15% sedangkan untuk rekomendasi tidak di utamakan di cluster ini karena presentasi yang sedikit, pemberian beasiswa bias dilakukan di cluster ini dikarenakan pendapatan orang tua yang paling sedikit dengan rata rata 1.200.000 dan membuka *stand* di beberapa kecamatan yang dominan. Untuk Hubungan masyarakat dan internet promosi juga dapat dijadikan strategi karena bersifat global.

1. Strategi Promosi berdasarkan Media Periklanan

   a. Brosur

   Salah satu media promosi yang digunakan adalah dengan menggunakan brosur. Kelebihan dari brosur ini adalah dapat menampilkan informasi baik secara visual maupun secara verbal. Sehingga sebuah brosur bisa menyampaikan informasi yang jelas sehingga konsumen bias langsung memahami maksud yang disampaikan serta brosur juga dapat dibuat sesaui keinginan karakter perusahaan.

   Brosur merupakan salah satu jenis media cetak untuk mempromosikan atau mengiklankan produk atau jasa dari suatu perusahaan. Bisa dikatakan bahwa beriklan dengan menggunakan brosur merupakan salah satu cara paling tradisional dalam teknik pemasaran. Dalam strategi ini diambil dari atribut media dan asal sekolah, dimana brosur lebih diutamakan data dibawah merupakan hasil dari analisis data mining didapat bahwa brosur lebih di utamakan di sekolah-sekolah pada Tabel 4.25.

   Tabel 4.25 Nama Asal Sekolah untuk Brosur

Asal Sekolah Berdasarkan Brosur		
SMPN 1 Limbangan	SMPN 1 Nagreg	MTS Alhidayah Ciaro
SMPN 2 Limbangan	SMPN 3 Kadungora	MTS Fatih Bandung
SMPN 3 Limbangan	SMP PGRI Cicalengka	MTS Al Furqon
SMPN 4 Limbangan	SMPN 1 Cicalengka	MTS Mathalul Anwar

Tabel 4.26 merupakan tabel program studi yang paling banyak diminati dari media brosur.

Pada strategi promosi ini didapat dari proses data mining yang telah di analisis bahwa jurusan Rekayasa Perangkat Lunak (RPL) merupakan jurusan yang banyak diminati pada media Brosur dengan Presentasi 64,50

Tabel 4.26 Jurusan Banyak Diminati Pada Media Brosur

Program Studi Banyak Diminati Berdasarkan Brosur
Rekayasa Perangkat Lunak

Tabel 4.27 Sedangkan Bisnis daring dan pemasaran lebih sedikit dengan Presentasi 36, 50 % merupakan tabel Jurusan yang paling sedikit diminati dari media brosur.

Tabel 4.27 Jurusan Kurang Diminati Pada Media Brosur

Program Studi Kurang Diminati Berdasarkan Media Brosur
Bisnis Daring Dan Pemasaran

b. Spanduk

Spanduk sering dijadikan sebagai media iklan bisnis karena dirasa memiliki beberapa kelebihan yang bisa digunakan oleh para pebisnis atau pengusaha sebagai media iklan untuk mempromosikan suatu usaha. Salah satu contohnya media spanduk dijadikan sebagai media promosi sekolah. Dalam strategi ini diambil dari atribut media dan kecamatan karena sekolah yang terletak strategis diantara perbatasan dari beberapa kecamatan, kecamatan ini merupakan data calon siswa yang banyak mengatahui promosi melalui media spanduk dimana media Spanduk lebih diutamakan di kecamatan-kecamatan pada Tabel 4.28.

Tabel 4.28 Nama Kecamatan Untuk Media Spanduk

Kecamatan Berdasarkan Media Spanduk	
Nagreg	Rancaekek
Kadungora	Cikancung
BL. Limbangan	Cicalengka

Tabel 4.29 merupakan tabel Jurusan yang banyak diminati dari media spanduk

Pada strategi promosi ini, didapat dari proses data mining yang telah di analisis bahwa jurusan Bisnis Daring Dan Pemasaran (BDP) merupakan jurusan yang banyak diminati pada media spanduk dengan presentasi 52,30 %

Tabel 4.29 Nama Program Studi Banyak Diminati pada Media Spanduk

Program Studi Diminati Berdasarkan Media Spanduk
Bisnis Daring dan pemasaran

Sedangkan jurusan Rekayasa Perangkat Lunak (RPL) lebih sedikit dengan Presentasi 48,70 % merupakan tabel Jurusan yang paling sedikit diminati dari media spanduk.

Tabel 4.30 merupakan tabel program studi yang kurang diminati dari media spanduk.

Tabel 4.30 Nama Program Studi Kurang Diminati pada Media Spanduk

Program Studi Kurang Diminati Berdasarkan Media Spanduk
Rekayasa Perangkat Lunak

2. Strategi Promosi Berdasarkan Penjualan Personal

a. Rekomendasi

Rekomendasi merupakan salah satu media promosi dimana calon siswa dipilih berdasarkan rekomendasi yang diajukan oleh sekolah asal sesuai kriteria ataupun kriteria pemilihan oleh sekolah langsung kriteria yang dipilih adalah siswa yang kurang mampu, siswa yatim, siswa yatim piatu dan berprestasi dibidang akademik. berdasarkan Strategi ini diambil dari atribut media dan asal sekolah, dimana rekomendasi lebih banyak dari sekolah-sekolah pada Tabel 4.31.

Tabel 4.31 Nama Sekolah untuk Rekomendasi

Asal Sekolah Berdasarkan Rekomendasi	
MTS Alhidayah Ciaro	SMPN 3 Limbangan
MTS Fatih Bandung	SMPN 4 Limbangan
MTS Al Furqon	SMPN 1 Nagreg
SMP Maarif Cicalengka	SMPN 3 Kadungora
SMP Ganjar Sabar	SMPN 1 Cicalengka

Pada strategi promosi ini, didapat dari proses data mining yang telah di analisis bahwa jurusan Bisnis Daring Dan Pemasaran (BDP) merupakan jurusan yang banyak diminati pada media rekomendasi dengan presentasi 61,20 %

Tabel 4.32 merupakan tabel program studi yang paling banyak diminati dari media rekomendasi.

Tabel 4.32 Tabel Nama Program Studi Banyak Diminati pada Media Rekomendasi

Program Studi Banyak Diminati Berdasarkan Rekomendasi
Bisnis Daring Dan Pemasaran

Sedangkan jurusan Rekayasa Perangkat Lunak (RPL) lebih sedikit dengan Presentasi 39,80 % merupakan tabel Jurusan yang paling sedikit diminati dari media rekomendasi.

Tabel 4.33 merupakan tabel program studi yang kurang diminati dari media rekomendasi.

Tabel 4.33 Tabel Nama Program Studi Kurang Diminati pada Media Rekomendasi

Program Studi Sedikit Diminati Berdasarkan Rekomendasi
Rekayasa Perangkat Lunak

b. Presentasi Sekolah

Presentasi sekolah merupakan media promosi dimana proses promosi dilakukan dengan cara berhadapan langsung dengan siswa siswi SMP dimana dalam proses presentasi ini tim promosi berhadapan dengan siswa siswi SMP sehingga dimana ketika ada siswa ada yang bertanya maka tim promosi akan langsung menjawab sehingga promosi menggunakan presentasi lebih efisien. Strategi ini diambil dari atribut media dan asal sekolah. Berdasarkan hasil wawancara, presentasi sekolah dilakukan pada sekolah-sekolah yang akan dibagikan brosur. Sehingga nama asal sekolah sama dengan Tabel 4.31.

3. Strategi promosi berdasarkan promosi penjualan

a. Beasiswa pelajar dan diskon

Beasiswa pelajar dan diskon merupakan salah satu strategi promosi yang dijadikan bahan pertimbangan dikarenakan banyaknya siswa yang ingin bersekolah akan

tetapi kurangnya biaya sehingga perlu adanya program khusus upaya dalam memberikan Pendidikan yang layak untuk siswa yang kurang mampu dan untuk siswa yang berprestasi di berikan keringanan dalam biaya sekolahnya. Berdasarkan data di lapangan ada banyak sekali siswa yang kurang mampu Strategi ini diambil dari atribut media dan asal sekolah. Berdasarkan hasil wawancara, beasiswa pelajar dan diskon dilakukan pada sekolah-sekolah yang akan dibagikan brosur. Sehingga nama asal sekolah sama dengan Tabel 4.31.

4. Strategi promosi berdasarkan Hubungan Masyarakat

a. Membuka Stand di Kecamatan

Strategi ini diambil dari atribut program studi, gaji orang tua dan kecamatan. Berdasarkan hasil wawancara, membuka stand dilakukan pada kecamatan-kecamatan per program studi yang mempunyai gaji rata-rata menengah kebawah seperti Tabel 4.34.

Tabel 4.34 Kecamatan Berdasarkan Rata-Rata Gaji Menengah dan Rendah

Kecamatan Berdasarkan Gaji Menengah dan Rendah	
Limbangan	Leles
Selaawi	Cangkuang
Kadungora	Nagreg
Cicalengka	Cimanggung

b. Kerjasama Sekolah

Kerjasama sekolah merupakan salah satu upaya dalam strategi promosi sekolah dimana pihak sekolah melakukan kerjasama secara langsung dengan sekolah sasaran dalam kasus ini dimana sekolah sasaran adalah sekolah SMP yang berada dikawasan terdekat sekolah rekomendasi dan luar wilayah kecamatan dengan sekolah, kerjasama sekolah juga memudahkan pihak sekolah dalam mempromosikan sekolah karena sudah ada kerjasama yang telah disepakati sehingga mempermudah dalam mempromosikan sekolah. Strategi ini diambil dari atribut media dan asal sekolah. Berdasarkan hasil wawancara, bekerja sama dengan sekolah dilakukan dengan sekolah-sekolah yang banyak memberikan lulusannya ke SMK Bhakti Putra.

5. Strategi promosi berdasarkan pemasaran langsung

Strategi promosi berdasarkan pemasaran langsung dengan media internet dapat dilakukan untuk semua kelompok karena sifatnya global, manfaat Mempublikasikan kegiatan sekolah melalui media online Tidak hanya bagi pihak sekolah itu sendiri, namun juga bagi masyarakat luas yang butuh informasi lebih tentang lembaga dan dunia pendidikan salah satu manfaat mengapa strategi promosi internet di jadikan bahan pertimbangan dalam media promosi yang efektif dan efisien adalah untuk Mengenalkan Sekolah, Sekolah yang berkembang tentu ingin lebih dikenal masyarakat luas, sehingga perlu media promosi yang memiliki akses luas pula. Dengan publikasi online, hal itu dapat dicapai. Promosi profile sekolah, kegiatan, prestasi, dan kemajuan pendidikan di suatu sekolah akan menarik minat calon siswa mendaftar. Mendokumentasikan Kegiatan Sekolah, Mendokumentasikan suatu kegiatan atau hasil karya sangat butuh yang tahan lama. Bila hanya kumpulan foto-foto di album sekolah, akan lapuk termakan usia. begitu juga dengan hasil karya siswa yang dipajang, semakin lama semakin membutuhkan tempat yang lebih luas. Dengan didokumentasikan dan dipublikasikan lewat media online, utamanya website, tentu tidak akan hilang atau lapuk. Bahkan bisa dilihat meski sudah beberapa tahun berlalu. Meningkatkan Budaya Literasi, Kegiatan yang di-publish lewat media online, menarik minat siswa atau warga sekolah mengunjungi media tersebut. Terlebih di website yang menjadi rekanan sekolah itu. Dengan demikian, budaya literasi akan semakin meningkat dilakukan pihak sekolah yang bersangkutan. Ajang Kompetisi Melalui media online, warga sekolah dapat mengikuti berbagai ajang kompetisi yang diselenggarakan oleh berbagai lembaga atau instansi. Baik untuk kepala sekolah, guru, maupun para siswa. Bila bekerjasama dengan sebuah website pendidikan, paling tidak siswa akan berkompetisi dalam menulis artikel, kegiatan, opini, dan juga karya sastra. Mereka akan berlomba agar tulisannya dimuat di website tersebut.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

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

1. Setelah di analisis dan dilakukan proses perbandingan Algoritma *K-Means* menjadi algoritma yang lebih baik dibandingkan dengan *K-Medoids*. Pada penelitian ini, bisa dilihat berdasarkan dari nilai validasi dunn *Index*, dimana nilai dunn *Index* paling tinggi pada *K-Means* K=5 sebesar 0,988 dan K=4 sebesar 0,488 untuk *K-Medoids*. Selain itu, dalam proses pengolahannya *K-Means* hanya membutuhkan waktu rata-rata 1 detik sedangkan algoritma *K-Medoids* membutuhkan waktu rata-rata 1 menit 5 detik artinya apabila jumlah pengelompokkan yang ditentukan lebih besar, proses pengolahan datapun semakin lama.
2. Kajian data *Mining* profil calon siswa yang dihasilkan algoritma *K-Means* dapat memberikan sebuah strategi promosi berdasarkan *promotion* mix, dimana brosur, persentasi sekolah, dan pemberian beasiswa lebih di utamakan pada *cluster* 0, *cluster* 1 dan *cluster* 3, spanduk lebih di utamakan pada *cluster* 0 dan *cluster* 2, dan rekomendasi lebih di utamakan pada *cluster* 2 dan *cluster* 4. Membuka stand di kecamatan lebih diutamakan pada *cluster* 0, *cluster* 1, *cluster* 2 dan *cluster* 3. Mengadakan perlombaan dan kerjasama dapat dilakukan pada semua *cluster.*
3. Profil calon siswa yang dihasilkan algoritma *K-Means* sebanyak 5 *cluster* (K=5), dimana *cluster* 0 dan *cluster* 1 merupakan kelompok calon siswa yang mempunyai rata-rata gaji orang tua rendah, *cluster* 2 dan 3 merupakan *cluster* yang mempunyai rata-rata gaji orang tua sedang, sedangkan *cluster* 4 merupakan *cluster* yang mempunyai rata-rata gaji orang tua tinggi

5.2 Saran

Adapun saran dari penelitian diantaranya :

1. Di penelitian selanjutnya agar bisa membandingkan dengan menggunakan algoritma *cluster*ing lainnya.
2. Menambahkan beberapa atribut lain yang tidak digunakan pada penelitian ini sehingga dapat menghasilkan strategi promosi sekolah yang lebih baik lagi.
3. Menambah jumlah data yang digunakan sebagai sumber data agar hasil nya lebih baik lagi.

DAFTAR PUSTAKA

Anggraeni, N. L. (2019). TEKNIK CLUSTERING DENGAN ALGORITMA K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG. *Jurnal Teknologi Informasi dan Pendidikan*, Vol. 12, No. 2.

Chasanah, T. T. (2017). PENENTUAN STRATEGI PROMOSI PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA CLUSTERING K-MEANS. *IC-Tech*, 39-44.

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.

Chowdary, N. S. (2014). Evaluating and Analyzing Clustersin Data Miningusing Different Algorithms. *IJCSMC*, 86-99.

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 ALGORITHM TO INCREASE SCALABILITY AND EFFICIENCY FOR LARGE DATASET. *IJRET*, 150-153.

Guoil, L. d. (2013). The Improved Research on K-Means Clustering Algorithm in Initial Values. International Conference on Mechatronic Science Electric Enginering and Computer (MEC).

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

Klinkenberg, M. H. (2013). *Rapidminer. Data Mining Use Cases and Business Analitics Aplications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series).* USA: CRC Press.

Kurniawati, I. (2017). Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru. *IKRAITH-INFORMATIKA, Vol. 1, No. 2*, 70-79.

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*, 911-916.

Maimon, 0. (2010). *Data Mining and Knowledge Discovery Handbook.* 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.04, No.01.

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 MAHASISSWA BARU UNIVERSITAS SATYA NEGARA INDONDESIA [ALGORITHMA K-MEANS CLUSTERING STRATEGY MARKETING ADMISSION UNIVERSITAS SATYA NEGARA INDONESIA]. *Seminar Nasional Cendekiawan*, 685-690.

Surmaningtyas, N. (2015). *Merancang sistem strategi promosi sekolah dengan metode k-means clustering di smkn 1 ngasem Kediri.* Kediri: simki.unpkediri.ac.id.

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*, 54-65.

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.

DAFTAR LAMPIRAN