



## Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari)

Widhi Lestari

STMIK Bina Bangsa Kendari  
widhi.nanga2@gmail.com

### Abstrak

Data mahasiswa STMIK Bina Bangsa Kendari setiap tahunnya terus bertambah dan menghasilkan data yang berlimpah sehingga terjadi penumpukan data. Data yang berlimpah perlu di lakukan pengolahan data untuk menggali informasi yang terdapat didalam data tersebut. Tujuan penelitian ini untuk mengkluster data mahasiswa di STMIK Bina Bangsa Kendari melalui proses data mining dengan menggunakan algoritma K-Means untuk pembentukan cluster. Atribut data digunakan adalah nama mahasiswa, jurusan, sekolah asal, nilai IPK. Data yang digunakan adalah data mahasiswa angkatan 2018/2019 dengan sampel data 171 items dan sumber data berasal dari bagian akademik dan admisi STMIK Bina Bangsa Kendari. Cluster mahasiswa yang terbentuk ada 2 yaitu Cluster 1 (cluster\_0) 72 items, dan Cluster 2 (cluster\_1) 99 items. Hasil dari penelitian ini di gunakan sebagai dasar untuk menentukan strategi promosi program studi yang ada di STMIK Bina Bangsa Kendari berdasarkan hasil cluster program studi yang banyak di minati dari masing-masing sekolah.

**Kata Kunci:** Data Mining, K-Means, Clustering, Promosi.

### Abstract

*The data of the students of the STMIK Bina Bangsa Kendari every year continues to grow and produce abundant data, causing data to accumulate. Abundant data needs to be processed by data to dig up the information contained in the data. The purpose of this study was to cluster student data at STMIK Bina Bangsa Kendari through a data mining process using the K-Means algorithm for cluster formation. The data attributes used are the name of the student, department, school of origin, GPA. The data used is the data of students in the class of 2018/2019 with a data sample of 171 items and the data source comes from the academic and admission department of the STMIK Bina Bangsa Kendari. There were 3 student clusters, namely Cluster 1 (cluster\_0) 72 items, and Cluster 2 (cluster\_1) 99 items. The results of this study are used as a basis for determining the study program promotion strategies in STMIK Bina Bangsa Kendari based on the results of the study program clusters that are in great interest from each school.*

**Keywords:** Data Mining, K-Means, Clustering, Promotion.

### 1. Pendahuluan

STMIK Bina Bangsa Kendari merupakan salah satu perguruan tinggi swasta yang berada di wilayah Sulawesi Tenggara. Mahasiswa tersebut berasal dari berbagai daerah di Sulawesi Tenggara di antaranya Kota Kendari, Konawe, Konawe Selatan, Konawe Kepulauan, Konawe Utara, Bombana, Muna. Selain itu, banyak pula berasal dari luar Sulawesi Tenggara seperti, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat.

Penerimaan mahasiswa baru yang di lakukan setiap tahun menghasilkan data yang berlimpah berupa profil mahasiswa baru. Berdasarkan berlimpahnya data mahasiswa, terdapat informasi yang dapat di ketahui dengan cara melakukan pengolahan data. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru.

Informasi tersebut dapat di gunakan sebagai penunjang dalam pengambilan keputusan bagi institusi.

Pengolahan data berukuran besar yang mempunyai jumlah *record* dan jumlah atribut yang cukup banyak tidak dapat dilakukan dengan mudah. Salah satu cara yang dapat diterapkan untuk mengolah data yang berukuran besar adalah dengan menerapkan penggunaan Data Mining. Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi. Karena di dalam data mining sendiri terdapat cara dan teknik dalam pemenuhan kebutuhan salah satunya adalah kebutuhan informasi yang luas, dan dari informasi yang kita dapat bisa kita gunakan sebagai suatu keputusan atau menentukan sebuah kualitas dalam menentukan suatu keputusan. Adapun kumpulan dari sebuah data atau informasi memiliki banyak potensi untuk dijadikan suatu kesimpulan dalam mengambil sebuah keputusan dengan melakukan analisis dan menggali suatu informasi yang terdapat di dalam sebuah data.

Dari definisi Data Mining yang luas terdapat banyak jenis teknik analisa yang dapat digolongkan dalam Data Mining. Teknik Data Mining yang akan digunakan dalam penelitian ini adalah *Clustering*. *Clustering* merupakan salah satu teknik data mining yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam *cluster (group)* sehingga setiap cluster akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya.

Penelitian dengan tema analisis data mahasiswa dengan data mining untuk mendukung strategi promosi perguruan ini sebelumnya sudah pernah dilakukan, tetapi tempat dan algoritma yang di gunakan berbeda-beda. Adapun penelitian sebelumnya yang digunakan untuk dapat dijadikan bahan pertimbangan dan di harapkan dapat membantu dalam penelitian ini dan penelitian berikutnya. Penelitian [1] menggunakan metode *Association Rules* dalam menganalisa hubungan daerah dan minat mahasiswa terhadap suatu prodi pada Universitas Nusantara PGRI Kediri yang digunakan oleh pengambil keputusan sehingga dapat memberikan rekomendasi wilayah promosi. Peneliti [2] menggunakan metode *clustering* data penduduk miskin dengan penerapan algoritma K-Means untuk membantu dalam pengelompokan data keluarga miskin agar bantuan dapat tersalurkan dengan tepat. Peneliti [3] pengelompokan data mahasiswa berdasarkan nilai dengan menggunakan algoritma K-Means dan EM (*Expectation Maximization*) untuk melihat sekolah mana yang berpotensi menghasilkan lulusan dengan nilai yang baik.

Berdasarkan pertimbangan dari uraian tersebut diatas dan berdasarkan referensi dari penilitan sebelumnya maka teknik data mining sangat di perlukan untuk mengolah data mahasiswa agar dapat mengetahui bagaimana hubungan daerah dan minat mahasiswa terhadap suatu program studi STMIK Bina Bangsa Kendari sehingga dapat memberikan rekomendasi strategi wilayah promosi.

Tujuan pengolahan data mahasiswa STMIK Bina Bangsa Kendari yaitu menerapkan algoritma K-Means Clustering pada data penerimaan mahasiswa baru STMIK Bina Bangsa Kendari. Atribut data yang akan digunakan adalah nama mahasiswa, jurusan, sekolah asal, nilai IPK. Data yang digunakan adalah data mahasiswa angkatan 2018/2019 dan sumber data berasal dari bagian akademik dan admisi STMIK Bina Bangsa Kendari. Pengolahan data menggunakan teknik data mining dengan algoritma K-Means *clustering*. Manfaat yang diharapkan adalah membantu bagian promosi STMIK Bina Bangsa Kendari dalam melakukan promosi dan mencari calon mahasiswa baru.

## 2. Metode

### 2.1. Data Mining

Data Mining adalah proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi di dalam timbunan data yang banyak. Menurut Subekti

Mujiasih dimuat dalam [4], Data Mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran Data Mining ini bisa dipakai untuk membantu pengambilan keputusan di masa depan. Pengembangan KDD ini menyebabkan penggunaan pattern recognition semakin berkurang karena telah menjadi bagian Data Mining.

## 2.2. Interpretation / Evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

## 2.3. Clustering

Salah satu teknik yang dikenal dalam data mining yaitu clustering. Pengertian clustering keilmuan dalam data mining adalah pengelompokan sejumlah data atau objek ke dalam *cluster* (*group*) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model cluster dan menghitung jumlah cluster yang optimal sehingga dapat dihasilkan *cluster* yang paling baik. Ada dua metode *clustering* yang kita kenal, yaitu hierarchical clustering dan partitioning. Metode hierarchical clustering sendiri terdiri dari complete linkage clustering, single linkage clustering, average linkage clustering dan centroid linkage clustering. Sedangkan metode partitioning sendiri terdiri dari k-means dan fuzzy k-means [5]. Pengelompokan (*clustering*) merupakan bagian dari ilmu data mining yang bersifat tanpa arahan (*unsupervised*). *Clustering* adalah proses pembagian data ke dalam kelas atau cluster berdasarkan tingkat kesamaannya. Dalam *clustering*, data yang memiliki kesamaan dimasukkan ke dalam *cluster* yang sama, sedangkan data yang tidak memiliki kesamaan dimasukkan dalam *cluster* yang berbeda [6].

## 2.4. Algoritma K-Means

K-Means *Clustering* adalah, K dimaksudkan sebagai konstanta jumlah cluster yang diinginkan, Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai cluster, sehingga K-Means Clustering adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode K-Means berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain.

Berikut adalah rumus untuk menentukan jarak data dari masing-masing centroid :

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Keterangan :

D = titik dokumen

P = data *record*

Q = data *centroid*

## 2.5. Teknik Pengumpulan Data

### 1) Observasi

Metode Observasi merupakan suatu metode pengambilan data dengan cara mengamati langsung tempat dan objek yang akan diteliti. Dalam penelitian ini yang menjadi objek penelitian yaitu STMIK Bina Bangsa Kendari.

### 2) Studi Literatur

Studi ini dimaksudkan untuk mengumpulkan dan memperoleh data yang tepat untuk penelitian dengan cara mempelajari, membaca, dan mencatat literatur dari beberapa jurnal yang berkaitan dengan permasalahan di atas.

## 2.6. Teknik Analisis Data

Pada tahap ini penulis menggunakan Proses *Knowledge Discovery in Databases* (KDD). Proses tersebut terdiri dari:

### 1) Data Selection

Pada tahap ini penulis mengambil profil data mahasiswa dari pihak Akademik STMIK Bina Bangsa Kendari. Data mahasiswa yang di jadikan sampel sebanyak 171 data untuk mahasiswa baru tahun ajaran 2018/2019. Kemudian dipilih beberapa atribut sesuai dengan kebutuhan. Dari beberapa atribut yang ada dipilih 3 atribut penting sebagai bahan penelitian, yaitu program studi, asal sekolah, dan nilai IPK.

**Tabel 1. Data Mahasiswa Yang di Peroleh**

No	Nama Mahasiswa	Jurusan	Asal Sekolah	IPK
1	Aga Sampe Pasang	Sistem Informasi	SMAN 5 KENDARI	3.20
2	Novi	Sistem Informasi	SMAN 1 LAWA	3.55
3	Nur Eka Putri Yan Sari	Sistem Informasi	SMKN 1 KABANGKA	3.75
4	Rezma Oktavia Ramli	Sistem Informasi	SMA NEGERI 4 RAHA	3.85
5	Erni	Sistem Informasi	SMK SWASTA LAGHUMBO	3.20
6	Uci Cahyani	Sistem Informasi	SMAN 1 LEMBO	3.20
7	Helma Hastuti	Sistem Informasi	SMAN 1 WAWONII BARAT	1.90
8	Elisabeth Margareta Lelo	Sistem Informasi	SMAN 1 WAWONII TENGAH	3.30
9	Suherman	Sistem Informasi	SMAN 16 KONAWA SELATAN	0.40
10	Wa Ode Ibo	Sistem Informasi	SMA NEGERI 1 LAWA	2.45
11	Ayu Nurmayanti Amir	Sistem Informasi	SMA N 1 TONGAUNA	3.35
12	Siti Vadyan Latif	Sistem Informasi	SMK N 04 KENDARI	3.45
13	Gian Luigi Fernanda	Sistem Informasi	SMAN 5 KONSEL	3.60
14	Saif Akbar	Sistem Informasi	MAN 1 KENDARI	3.90
15	Gusman	Sistem Informasi	SMAN 9 KONAWA SELATAN	2.95
16	Chichy Anacitra	Sistem Informasi	SMAN 1 LASOLO	3.45
17	Ramadhan Putra Januar Saranani	Sistem Informasi	SMAN 1 WONGGEDUKU	2.94
18	Aviva Nur Fitriani	Sistem Informasi	SMAN 02 BOMBANA	3.55
19	Sri Wahyuni	Sistem Informasi	SMA 1 WAWOTOB	3.30
20	Sulifita Sari	Sistem Informasi	SMKN 1 KENDARI	3.15

21	La Ode Mpoaro	Sistem Informasi	SMAN 1 MUNA	3.30
22	Iin Santia	Sistem Informasi	SMAN 4 KONSEL	3.30
23	Windsari	Sistem Informasi	SMAN 1 WAWONII TIMUR	3.25
24	Wd. Enggi Rahmadani	Sistem Informasi	SMA NEGERI 1 KOLINSUSU	2.05
25	Nur Fatma Sari	Sistem Informasi	SMAN 1 KONAWA	3.10
26	Abdul Fajar	Sistem Informasi	SMAN 1 WAWOTOBI	3.45
27	Bintang	Sistem Informasi	SMAN 19 KONSEL	3.45
28	Alda	Sistem Informasi	SMA NEGERI 1 UNAAHA	3.35
29	LIA ANJELISTA	Sistem Informasi	SMKN 4 KENDARI	3.50
30	Mardin	Sistem Informasi	SMK N 4 KENDARI	3.35
31	Ebi Kusbianto	Sistem Informasi	SMAN 1 WAWONII UTARA	2.55
32	Gabrielia Larysa Pongmanda	Sistem Informasi	SMAN 1 UNAAHA	3.65
33	ISMAIL	Sistem Informasi	SMKN 1 UNAAHA	0.40
34	Oci Astuti	Sistem Informasi	SMAN 1 WAWONII TENGGARA	3.20
35	Andi Riswan	Sistem Informasi	SMAN 3 WANGI WANGI	3.20
36	Muh. Syafril	Sistem Informasi	SMAN 1 DEKAI	2.10
37	Lilis Pratiwi	Sistem Informasi	SMKN 1 KULISUSU UTARA	3.00
38	Rizki Mila Dewi	Sistem Informasi	SMA N 2 KULISUSU	3.20
39	Komang Waraasta	Sistem Informasi	SMKN 1 UNAAHA	2.95
40	Unaldin Muhlis	Sistem Informasi	SMAN 1 ANGGABERI	3.20
41	ISMAIL	Sistem Informasi	SMAN 1 RUMBIA	2.94
42	Muh. Adi Muryanto	Sistem Informasi	SMAN 1 KULISUSU	3.10
43	Hebriyanto	Sistem Informasi	SMAN 1 ASERA	2.60
44	Arling Marusing	Sistem Informasi	MA MUTMAINNAH LIMBO	3.20
45	Widya Jafir	Sistem Informasi	MA MUTMAINNAH LIMBO	3.30
46	Munaya	Sistem Informasi	SMA N 1 KULISUSU	3.40
47	Wihan Pradana Ode	Sistem Informasi	SMAS BINA BANGSA KENDARI	2.40
48	Alan Tedisa	Sistem Informasi	SMAN 1 KULISUSU UTARA	2.75
49	Heri Pebrianto Alam	Sistem Informasi	MAN 1 KENDARI	3.90
50	YEYEN PRATIWI	Sistem Informasi	SMK N 4 KENDARI	3.50
51	Rahmad Akbar	Sistem Informasi	SMAN 6 KONSEL	1.60
52	AINUN ILMIH	Sistem Informasi	SMAN 1 LEMBO	3.20
53	PANDI ALFAYET	Sistem Informasi	SMAN 1 WAWONII UTARA	2.40
54	Lili Novrianti	Sistem Komputer	SMAN 1 LAWAN	3.45
55	Wa Ria	Sistem Komputer	SMAN 1 TONGKUNO	3.36
56	King Alfred Torino	Sistem Komputer	SMA	1.91
57	Elda	Sistem Komputer	SMAN 1 UNAAHA	3.55
58	Umar Suwandi	Sistem Komputer	SMA N 1 POLI-POLIA	3.64
59	Idul Rizki Moita	Sistem Komputer	SMKN 1 UNAAHA	3.64
60	La Ode Muhammad Hidayat Al Khaliq	Sistem Komputer	SMK4 KENDARI	3.14
61	Nur Vadila Cahya Lestari	Sistem Komputer	SMA NASIONAL WAWOTOBI	3.14

62	Rizal Basri	Sistem Komputer	SMAN 10 KENDARI	2.64
63	Sri Utami	Sistem Komputer	SMAN 1 TOBADAK	3.18
64	Olvy Yolanda	Sistem Komputer	SMANNEGERI 3 SIGI	3.36
65	Karina	Sistem Komputer	SMA N 11 KONAWA SELATAN	3.27
66	Armin	Sistem Komputer	SMAN 1 OHEO	3.36
67	Wa Jupulina	Sistem Komputer	SMAN 2 BAUBAU	3.36
68	Syawal	Sistem Komputer	SMAN 1 WIWIRANO	2.86
69	Indarmawati Abd.Salam	Sistem Komputer	SMAN 1 BUNGKU TENGAH	3.36
70	Putri. T. Firdaus	Sistem Komputer	SMA KARTIKA XX-2 KENDARI	3.55
71	Murniati	Sistem Komputer	SMA N 1 LAWA	3.55
72	Igo Ilham	Sistem Komputer	SMAN 1 LASOLO	3.45
73	Sandy Saputra	Sistem Komputer	SMAN 1 LASOLO	3.45
74	Risal Saputra	Sistem Komputer	SMA NEGERI 1 PURIALA	4.00
75	Nusardin	Sistem Komputer	PKBM ANALIPU	2.55
76	Febri Esawaldi	Sistem Komputer	SMAN 1 ANGGABERI	3.00
77	Arifudin Hadi Putra	Sistem Komputer	SMKS KESEHATAN UNAAHA	2.91
78	Syawal	Sistem Komputer	SMK N 1 KABANGKA	2.50
79	Hesti Kumala	Sistem Komputer	SMAN 1 WAWONII TENGAH	3.36
80	Siti Muliana Sari	Sistem Komputer	SMKN 3 RAHA	3.36
81	Sahrullah	Sistem Komputer	SMAN 1 POLI-POLIA	3.50
82	Tria Saputri S. Samaga	Sistem Komputer	SMA N 7 KONSEL	3.36
83	Muh. Sigit Sopyan. M	Sistem Komputer	SMAN 1 PURIALA	3.27
84	La Ode Naalu	Sistem Komputer	SMA NEGERI 8 KENDARI	2.09
85	Ade Fikra	Sistem Komputer	SMA 1 LAMBUYA	2.45
86	Eko Pratama	Sistem Komputer	SMAN 8 KENDARI	1.64
87	Nurul Izmi	Sistem Komputer	SMKN 5 KONAWA SELATAN	3.36
88	Arwan	Sistem Komputer	SMAN 2 KENDARI	2.95
89	Peni Indriani	Sistem Komputer	SMAN 4 KONSEL	3.18
90	Rabbina. S	Sistem Komputer	SMKN 1 KENDARI	2.77
91	Sriana	Sistem Komputer	SMAN 1 LAMBUYA	3.27
92	LA ODE HARYONO	Sistem Komputer	SMK BINA BANGSA	2.82
93	Agustina	Sistem Komputer	SMAN 04 KONAWA SELATAN	3.27
94	Dwiki Ardiawan	Sistem Komputer	SMKN 7 KONAWA SELATAN	2.64
95	Sahrul	Sistem Komputer	SMAN 1 WAWONII BARAT	2.77
96	Muhammad Fahrin	Sistem Komputer	SMAN 1 WAWONII	1.91
97	Nurmila	Sistem Komputer	SMA NEGERI 1 ASERA	3.36
98	PERLINDA APRESIA	Sistem Komputer	SMAN 1 LASOLO	2.77
99	Anita Wulandari	Sistem Komputer	SMA N 7 KONSEL	3.36
100	Lilis Tiawati	Sistem Komputer	SMAN 6 BOMBANA	3.55
101	Feti Pajeriati	Sistem Komputer	SMAN 1 WANUT	2.86
102	Muh Rajab Akbar	Sistem Komputer	SMA NEGERI 6 KENDARI	3.45

103	Nursam	Sistem Komputer	SMAN 1 WAWONII UTARA	2.45
104	Hendrik	Sistem Komputer	SMAN 2 ASERA	3.18
105	Armin	Sistem Komputer	SMKN 1 KONTUNAGA	2.73
106	Peti Pebriani	Sistem Komputer	SMA 5 KONSEL	2.36
107	Cahyani Rahmandani	Sistem Komputer	SMA NEGERI 1 SOROPIA	3.36
108	Siti Nurhalijah	Sistem Komputer	SMAN 1 KONAWA	3.36
109	Harianti	Sistem Komputer	SMAN 1 SAMPARA	3.27
110	Suhardin S. Rahman	Sistem Komputer	SMKN 1 BANGGAI	2.27
111	Ahmad Fahrian Angrayuda	Sistem Komputer	SMAN 2 UNAAHA	2.82
112	Citra Ramadani	Sistem Komputer	SMA 1 LABAKKANG	3.27
113	Sitti Nur Hafni	Sistem Komputer	SMAN 1 BONDOALA	1.55
114	Muh. Sulham	Sistem Komputer	SMAN 1 LASOLO	2.86
115	Herlita	Sistem Komputer	SMAN 1 BONDOALA	3.45
116	Idlan Abdullah	Sistem Komputer	SMKN 2 BOMBANA	3.23
117	Khaider Ali	Sistem Komputer	SMAN 1 WAWONII	0.41
118	Dhelfy Mulnazhary Nakir	Sistem Komputer	SMAN 1 KONSEL	3.36
119	Dina Lisnawati	Sistem Komputer	SMA satria	3.14
120	Asti Fania	Sistem Komputer	SMKN 1 KULISUSU UTARA	2.59
121	AGUNG CAHYONO	Sistem Komputer	SMAN 1 WAWOTOBI	1.82
122	Yusna	Sistem Komputer	SMKN 1 ANDOWIA	3.55
123	Fadel Muhammad	Sistem Komputer	SMAN 3 KENDARI	3.27
124	Riska Oktaviani	Sistem Komputer	SMAN 1 WAWOTOBI	2.73
125	Vebhy Febrima M	Sistem Komputer	SMAN 1 WAWOTOBI	3.18
126	Indra Hidayat	Sistem Komputer	SMAN 1 WAWONII BARAT	3.18
127	Yunita	Sistem Komputer	SMAN 1 WIWIRANO	3.18
128	Indriyani	Sistem Komputer	SMAN 1 OHEO	2.68
129	Lukman	Sistem Komputer	SMAN 1 LAWA	2.55
130	Hendra Tibowo	Sistem Komputer	SMA N 2 KUSAMBI	3.05
131	Muh. Syafriawan. B	Sistem Komputer	SMA 1 SOROPIA	2.55
132	Izal	Sistem Komputer	SAMN 1 LAWA	1.91
133	La Ode Arfan	Sistem Komputer	SMAN 1 LAWA	2.41
134	Muhammad Ade Novrianto	Sistem Komputer	SMKN 2 KENDARI	3.59
135	Gustinawati	Sistem Komputer	SMA N 1 LAWA	3.27
136	La Ode Andi Umar	Sistem Komputer	SMA N 1 LAWA	2.77
137	Sahrul Sihab	Sistem Komputer	SMAN 6 KENDARI	3.59
138	Liliana Putri	Sistem Komputer	SMAN 1 LASOLO	3.18
139	Firmansyah	Sistem Komputer	SMAN 2 ASERA	3.36
140	Pitno Brayatno	Sistem Komputer	SMKN 1 MOLAWA	3.41
141	Fajar	Sistem Komputer	SMAN 1 WIWIRANO	1.91
142	Aldin	Sistem Komputer	SMAN 1 WIWIRANO	2.77
143	Alfina	Sistem Komputer	SMKN 1 BUNGKU TIMUR	2.00
144	Sulfikar	Sistem Komputer	SMAN 1 LASOLO	2.27

145	Junaidin	Sistem Komputer	SMKN 1 MOLAWE	3.27
146	Ival Ma'Mun	Sistem Komputer	SMA KARTIKA XX-2 KENDARI	3.09
147	Cahaya Nur Fadillah	Sistem Komputer	SMKN 1 TIRAWUTA	3.68
148	Jecky	Sistem Komputer	SMK 02 KONSEL	0.36
149	Raslina	Sistem Komputer	SMK N 5 KONSEL	2.68
150	Razat	Sistem Komputer	SMAN 1 KULISUSU UTARA	3.18
151	Ayu Mustika	Sistem Komputer	SMA N 1 KULISUSU	3.27
152	Nengsih Ani Sapitri	Sistem Komputer	SMKN 1 MOLAWE	2.86
153	Rina Ningsih Sapitri	Sistem Komputer	SMKN 1 MOLAWE	2.86
154	Eti Mey Dyawati	Sistem Komputer	SMKN 1 MOLAWE	2.27
155	Izha Kiftia	Sistem Komputer	SMA	2.86
156	Andri Ardiansyah	Sistem Komputer	SMAN 1 LASOLO KEPULAUAN	2.05
157	Ardi Sasmita	Sistem Komputer	MA AL-MUHAJIRIN	3.68
158	Nining Rosmalita	Sistem Komputer	SMAN 1 WAWONII BARAT	1.55
159	Ajis	Sistem Komputer	SMA N 1 WAWONII BARAT	2.36
160	Muh. Afiful Hakim	Sistem Komputer	M.A ANNUR AZZUBAIDI	2.68
161	Yuli Melinda	Sistem Komputer	SMA 16 KONSEL	2.50
162	Adhi Firman	Sistem Komputer	SMKS AL MUJAHIDIN	3.68
163	Fitriatul Khoiriah	Sistem Komputer	MA ANNUR AZZUBAIDI	3.14
164	Hartina	Sistem Komputer	SMAN 1 WAWONII TIMUR	1.18
165	Masdar	Sistem Komputer	SMA N 1 TALIABU BARAT LAUT	2.36
166	Siti Syarah Harahap	Sistem Komputer	SMKN 3 KENDARI	2.68
167	Moh. Rohul Pramana	Sistem Komputer	SMA MUHAMMADIYAH 1 PALU	1.50
168	Jimi Rahmat	Sistem Komputer	SMA 1 LAMBANDIA	1.55
169	Ahmad Insan Kamil	Sistem Komputer	SMAN 1 KONAWE	1.45
170	Muhammad Miftahul Khairi Asis	Sistem Komputer	SMK NEGERI 2 KENDARI	2.86
171	Elva Damayana	Sistem Komputer	SMAN 1 KALEDUPA	3.18

### 1) *Pre-processing*

Sebelum proses clustering dilakukan, diperlukan *pre-processing* terlebih dahulu dengan cara menghapus data yang hilang, missing, atau duplikat pada atribut penting.

### 2) *Transformation*

Transformasi data dilakukan untuk mengubah data tujuannya adalah agar data dapat diolah dengan menggunakan metode *K-Means Clustering*. Adapun variabel yang digunakan pada data mahasiswa baru yaitu data Program Studi, Asal Sekolah dan Nilai IPK. Untuk data program studi dikelompokkan menjadi 2 kelompok yaitu program studi sistem informasi ditransformasikan dengan nilai 1 dan program studi sistem komputer ditransformasikan dengan nilai 2. Untuk asal sekolah SMA di transformasikan dengan nilai 1, asal sekolah SMK di transformasikan dengan nilai 2, dan asal sekolah selain SMA dan SMK di transformasikan dengan nilai 3. Untuk variable nilai IPK di kelompokkan menjadi 3 kelompok, yang pertama untuk nilai IPK dengan rata-rata  $\leq 2$  di transformasikan dengan nilai 1, nilai  $> 2$  ditransformasikan dengan nilai 2, dan nilai  $\geq 3$  ditransformasikan dengan nilai 3. Adapun hasil dari transformasi dapat di lihat pada table berikut ini.



**Tabel 2. Hasil Tranformasi Data**

ID	PRODI	ASAL SEKOLAH	Nilai IPK
K1	1	1	3
K2	1	1	3
K3	1	2	3
K4	1	1	3
K5	1	2	3
K6	1	1	3
K7	1	1	1
K8	1	1	3
K9	1	1	1
K10	1	1	2
K11	1	1	3
K12	1	2	3
K13	1	1	3
K14	1	3	3
K15	1	1	2
K16	1	1	3
K17	1	1	2
K18	1	1	3
K19	1	1	3
K20	1	2	3
K21	1	1	3
K22	1	1	3
K23	1	1	3
K24	1	1	2
K25	1	1	3
K26	1	1	3
K27	1	1	3
K28	1	1	3
K29	1	2	3
K30	1	2	3
K31	1	1	2
K32	1	1	3
K33	1	2	1
K34	1	1	3
K35	1	1	3
K36	1	1	2
K37	1	2	3
K38	1	1	3
K39	1	2	2
K40	1	1	3
K41	1	1	2
K42	1	1	3
K43	1	1	2
K44	1	3	3
K45	1	3	3
K46	1	1	3
K47	1	1	2
K48	1	1	2
K49	1	3	3
K50	1	2	3
K51	1	1	1
K52	1	1	3
K53	1	1	2
K54	2	1	3
K55	2	1	3
K56	2	1	1

K57	2	1	3
K58	2	1	3
K59	2	2	3
K60	2	2	3
K61	2	1	3
K62	2	1	2
K63	2	1	3
K64	2	1	3
K65	2	1	3
K66	2	1	3
K67	2	1	3
K68	2	1	2
K69	2	1	3
K70	2	1	3
K71	2	1	3
K72	2	1	3
K73	2	1	3
K74	2	1	3
K75	2	3	2
K76	2	1	3
K77	2	2	2
K78	2	2	2
K79	2	1	3
K80	2	2	3
K81	2	1	3
K82	2	1	3
K83	2	1	3
K84	2	1	2
K85	2	1	2
K86	2	1	1
K87	2	2	3
K88	2	1	2
K89	2	1	3
K90	2	2	2
K91	2	1	3
K92	2	2	2
K93	2	1	3
K94	2	2	2
K95	2	1	2
K96	2	1	1
K97	2	1	3
K98	2	1	2
K99	2	1	3
K100	2	1	3
K101	2	1	2
K102	2	1	3
K103	2	1	2
K104	2	1	3
K105	2	2	2
K106	2	1	2
K107	2	1	3
K108	2	1	3
K109	2	1	3
K110	2	2	2
K111	2	1	2
K112	2	1	3
K113	2	1	1
K114	2	1	2
K115	2	1	3
K116	2	2	3
K117	2	1	1
K118	2	1	3
K119	2	1	3
K120	2	2	2
K121	2	1	1

K122	2	2	3
K123	2	1	3
K124	2	1	2
K125	2	1	3
K126	2	1	3
K127	2	1	3
K128	2	1	2
K129	2	1	2
K130	2	1	3
K131	2	1	2
K132	2	1	1
K133	2	1	2
K134	2	2	3
K135	2	1	3
K136	2	1	2
K137	2	1	3
K138	2	1	3
K139	2	1	3
K140	2	2	3
K141	2	1	1
K142	2	1	2
K143	2	2	2
K144	2	1	2
K145	2	2	3
K146	2	1	3
K147	2	2	3
K148	2	2	1
K149	2	2	2
K150	2	1	3
K151	2	1	3
K152	2	2	2
K153	2	2	2
K154	2	2	2
K155	2	1	2
K156	2	1	2
K157	2	3	3
K158	2	1	1
K159	2	1	2
K160	2	3	2
K161	2	1	2
K162	2	2	3
K163	2	3	3
K164	2	1	1
K165	2	1	2
K166	2	2	2
K167	2	1	1
K168	2	1	1
K169	2	1	1
K170	2	2	2
K171	2	1	3

### 2.7. Data Mining Menggunakan Algoritma K-Means

Pada penelitian ini, penulis menggunakan teknik *clustering* dengan algoritma *K-means*. Berikut langkah-langkah algoritma *K-means*:

- a) Menentukan k sebagai jumlah kluster yang ingin di bentuk. Nilai K yang digunakan adalah 2.

**Tabel 3. Titik Pusat Cluster**

<b>Titik Pusat</b>			
Centroid 1	1	1	1
Centroid 2	2	1	3

- b) Membangkitkan nilai random untuk pusat kluster awal (*centroid*) sebanyak k.
- c) Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak *Euclidean Distance* hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan *Euclidian Distance*:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Dimana :

$x_i$  : data kriteria,

$\mu_j$  : centroid pada cluster ke-j

Sebagai contoh, akan dihitung jarak dari data mahasiswa pertama ke pusat cluster pertama dengan persamaan:

$$d(1,1) = \sqrt{(1-1)^2 + (1-1)^2 + (3-1)^2}$$

$$= 2$$

Jarak antara data mahasiswa baru pertama ke cluster pertama adalah 2.

$$d(1,2) = \sqrt{(1-2)^2 + (1-1)^2 + (3-3)^2}$$

$$= 1$$

Jarak antara data mahasiswa baru pertama ke cluster ke dua adalah 1.

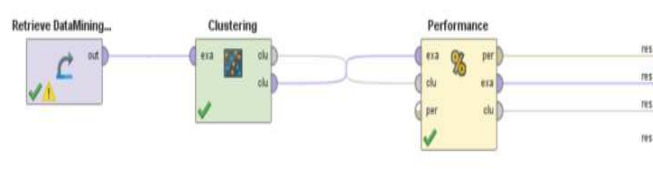
- d) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
- e) Melakukan perulangan dari langkah ke-2 hingga ke-4 sampai anggota tiap kluster tidak ada yang berubah.
- f) Jika langkah ke-5 telah terpenuhi, maka nilai pusat kluster ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

### 2.8. Interpretation/Evaluation

Pada tahap ini penentuan strategi promosi dari hasil *clustering* yang telah dilakukan. Strategi promosi tersebut sesuai dengan kebutuhan kampus dari hasil wawancara dan observasi.

### 3. Hasil dan Pembahasan

Hasil pengolahan data mahasiswa baru menggunakan algoritma K-Means *Clustering* dan *Software RapidMiner* dapat dilihat pada gambar berikut :



**Gambar 2. Pemodelan K-Means pada RapidMiner**

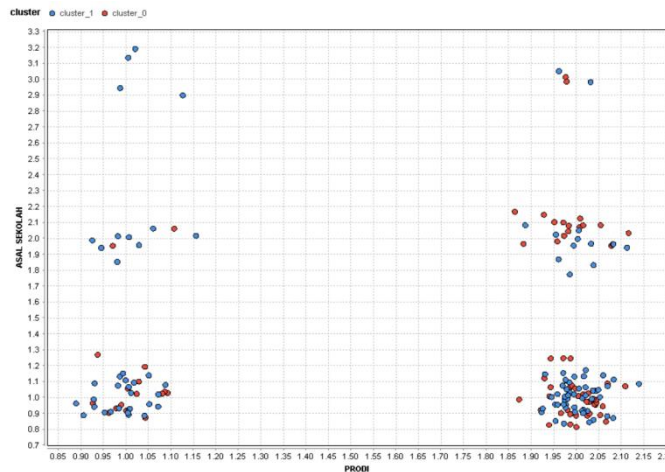
Pengolahan data mahasiswa baru menggunakan pemodelan K-Means *clustering* dengan jumlah data 171 dan jumlah clustering sebanyak 2 cluster, dengan mendefinisikan nilai k dengan jumlah cluster\_0 : 72 items , dan cluster 1 : 99 items.

```

Cluster Model
Cluster 0: 72 items
Cluster 1: 99 items
Total number of items: 171
    
```

**Gambar 3. Cluster Model**

Hasil penyebaran cluster 0, dan cluster 1 sebanyak 171 data pada pemodelan k-means clustering dengan menggunakan *rapidminer*, dan dapat di lihat pada gambar berikut :



**Gambar 4. Pemodelan K-Means Clustering pada RapidMiner**

Untuk kelompok data yang terdapat pada gambar diatas terdiri dari dua kelompok data. Kelompok pertama terlihat pada titik-titik penyebaran yang ditandai dengan warna hijau, kelompok kedua terlihat pada titik-titik penyebaran yang ditandai dengan warna biru, dan kelompok ketiga terlihat pada titik-titik penyebaran yang ditandai dengan warna merah. Adapun hasil analisa cluster pada gambar 3, berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antra titik pusat dengan data mahasiswa pada setiap atribut.

**Tabel 4. Hasil Analisa Cluster Satu (Cluster\_0)**

Hasil Cluster 1	
Program Studi :	Asal sekolah :
Sistem Informasi : 16	SMA = 14 SMK = 2 Lainnya = 0
Sistem Komputer : 56	SMA = 38 SMK = 16 Lainnya = 2
<b>Rata-Rata Nilai IPK : 2,30</b>	

**Tabel 5. Hasil Analisa Cluster Dua (Cluster\_1)**

Hasil Cluster 2	
Program Studi :	Asal sekolah :
Sistem Informasi : 37	SMA = 25 SMK = 8 Lainnya = 4
Sistem Komputer : 62	SMA = 49 SMK = 11 Lainnya = 2
<b>Rata-Rata Nilai IPK : 3,36</b>	

Dari data hasil *clustering* yang telah dilakukan di atas, maka dapat ditentukan beberapa strategi promosi dalam penerimaan mahasiswa baru yang dapat dilakukan oleh pihak STMIK Bina Bangsa Kendari dalam melakukan promosi program studi di masing-masing sekolah yang ada di provinsi Sulawesi Tenggara, sesuai dengan hasil yang paling banyak diminati berdasarkan dari masing-masing cluster yang terbentuk.

#### 4. Kesimpulan dan Saran

##### 4.1 Kesimpulan

Setelah melakukan tahapan penerapan algoritma k-means *clustering* ada beberapa kesimpulan yaitu : Penentuan *centroid* (titik pusat) pada tahap awal Algoritma K-Means sangat berpengaruh pada hasil *cluster* seperti pada hasil pengujian yang dilakukan dengan menggunakan 171 *dataset* dengan *centroid* yang berbeda menghasilkan hasil *cluster* yang berbeda juga. Setelah dilakukan pengelompokan data penerimaan mahasiswa menggunakan metode k-means *Clustering* terbentuk dua *cluster* yaitu cluster satu dengan jumlah 72 *items*, dan cluster dua dengan jumlah 99 *items*. Strategi promosi bagi calon mahasiswa baru nantinya akan mengikuti *cluster* yang terbentuk berdasarkan program studi yang paling banyak diminati di masing-masing sekolah.

##### 4.2 Saran

Adapun saran untuk penelitian berikutnya antara lain :

1. Pengelompokan data mahasiswa STMIK Bina Bangsa Kendari sebaiknya dilakukan setiap tahun ajaran baru.
2. Penelitian ini dapat dijadikan salah satu referensi untuk yang akan mendatang dalam melakukan promosi program studi yang diminati di masing-masing sekolah.
3. Pada penelitian berikutnya, diharapkan agar bisa di kombinasikan dengan menggunakan metode *clustering* lainnya untuk menghasilkan suatu penelitian yang lebih baik.

#### Daftar Pustaka

- [1] Elvira, A., Fana, W., Yogi, Y. (2015), "Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning", Jurnal Teknologi Informasi & Komunikasi Digital Zone, Volume 6, 2015.
- [2] Fajar, R., Risky A. (2017), "Penerapan Data Mining Menggunakan Association Rules untuk Mendukung Strategi Promosi Universitas Nusantara PGRI Kediri", SNATIKA, 2017.
- [3] Femi, D. (2017), "Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma Hard C-Means", Jurnal Ilmiah DASI, Volume 18, 2017.
- [4] Fadlina, (2014), "Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori", Majalah Ilmiah INTI, Vol. 3, No. 1, 2014.
- [5] Alfina, Santosa, Barkbah. "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Kedua dalam Cluster Data", Jurnal Teknik ITS, Vol. 1, No. 1, 2012.
- [6] Khotimah, T., Teknik, D. F., Studi, P., (2014). Pengelompokan Surat Dalam AL QUR ' AN Menggunakan Algoritma K-Means, Informatika, T., & Kudus, U. M, Vol. 5, No. 1, 2014.