

PENENTUAN STRATEGI MARKETING SEKOLAH MENENGAH KEJURUAN TERPADU LAMPANG SUBANG MENGUNAKAN METODE *K-MEANS CLUSTERING*

Rino Gupitha
Fakultas Ilmu Komputer, Universitas Subang

mbahhaya@gmail.com

Abstrak

Penentuan Strategi Marketing Sekolah Menengah Kejuruan Terpadu Lampung Subang Menggunakan Metode *K-Means Clustering*. Strategi penerimaan siswa baru yang dilakukan oleh setiap sekolah haruslah berbeda. Hal ini dapat diterapkan oleh sekolah agar dapat menjangkau para calon siswa untuk memilih sekolah tersebut. Penelitian ini bertujuan untuk mengelompokkan data calon siswa ke dalam sebuah *cluster*. Sehingga diperoleh kelompok berdasarkan dari kemiripan data, yang akan berada dalam satu *cluster*. Penelitian ini di aplikasikan dengan pendekatan *K-Means Clustering* yang dapat digunakan sebagai suatu acuan untuk menentukan strategi promosi yang lebih efektif dan efisien serta dapat meningkatkan jumlah pendaftaran siswa baru selanjutnya.

Kata kunci : Strategi, Data Mining, K-means Klaster

Abstract

Determination of Marketing Strategy of Vocational High Schools Terpadu Lampung Subang Using the K-Means Clustering Method. The new student admission strategy adopted by each school must be different. This can be applied by schools in order to attract prospective students to choose the school. This study aims to group prospective student data into a cluster. So that the group is obtained based on the similarity of data, which will be in one cluster. This research is applied with the *K-Means Clustering* approach which can be used as a reference to determine more effective and efficient promotion strategies and can increase the number of new student registrations subsequently.

Keywords: Strategy, Data Mining, K-means Cluster

Pendahuluan

Pendidikan dalam kehidupan suatu negara memegang peranan yang sangat penting untuk menjamin kelangsungan hidup negara dan bangsa. Setiap tahun ajaran baru, sekolah khususnya sekolah swasta berlomba – lomba untuk dapat menarik minat calon siswa. Dengan kehebatan promosi para lulusannya, dan menyebutkan bahwa lulusannya bisa langsung bekerja serta memiliki jabatan di instansi atau perusahaan maupun menjadi seorang wirausaha yang sukses. Sekolah sebagai institusi pendidikan telah memiliki data akademik dan administrasi dalam jumlah yang sangat besar, namun hanya sebagian kecil yang dimanfaatkan dari data tersebut.

Kemajuan teknologi informasi saat ini semakin berkembang pesat dan hampir mencakup segala bidang kehidupan. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang cukup besar, seperti data siswa aktif dan hasil dari proses belajar. Seperti halnya pada proses penerimaan siswa baru dalam sebuah sekolah akan

menghasilkan data – data baru berupa profil dari siswa baru tersebut. Tahap selanjutnya adalah siswa akan melakukan kegiatan pembelajaran di setiap semester, sehingga dapat diketahui data nilai siswa pada akhir semester. Hal ini akan terjadi secara berulang pada sebuah sekolah. Apabila dilakukan pengolahan pada kedua sumber data tersebut, maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi marketing penerimaan siswa baru tahun berikutnya.

Ada banyak cara yang dapat digunakan untuk menganalisis dan menentukan strategi marketing yang cocok dan tepat sasaran, salah satunya yaitu *data mining*. *Data mining* juga dapat digunakan sebagai prediksi untuk memperkirakan nilai masa mendatang.

Dengan menerapkan teknik ini sumber data akan di olah untuk kemungkinan sekolah tersebut banyak di pilih oleh calon siswa. Pengolahan tersebut dapat dilakukan menggunakan *data mining* dengan metode *K-Means Clustering*. Atribut yang dibutuhkan adalah tempat tinggal siswa, asal sekolah siswa, dan nilai siswa.

Penelitian ini dilakukan untuk pengelompokan data sehingga sekolah dapat melakukan strategi marketing yang tepat untuk mendapatkan calon siswa baru.

Kajian Teori

Penelitian Terkait

Ediyanto, dkk telah menggunakan metode *K-Means Cluster Analysis* yang cukup efektif diterapkan dalam proses pengklasifikasian karakteristik terhadap objek penelitian. Algoritma *K-Means* juga tidak terpengaruh terhadap urutan objek yang digunakan. Hal ini dibuktikan ketika mencoba menentukan secara acak titik awal pusat *cluster* dari salah satu objek pada permulaan perhitungan [1].

Ong Johan Oscar, penelitian ini menunjukkan bahwa hasil dari pengolahan data mahasiswa membantu pihak marketing *President University* dalam melakukan pemasaran dan mencari calon mahasiswa baru dari berbagai kota di Indonesia, dan hasilnya cukup efisien dan efektif [2].

T. Gomasathit melakukan penelitian untuk mengidentifikasi cakupan awan di atas wilayah Thailand dengan menggunakan data satelit dan algoritma *K-Means Clustering*. Namun, hasilnya dijelaskan dalam penelitian memberikan beberapa informasi awal tentang kemungkinan analisis cuaca dari perkiraan awan, dan akan berguna untuk studi lanjut ketika data lainnya tersedia [3].

Strategi Pemasaran dalam Lembaga Pendidikan

Kotler menjelaskan bahwa strategi pemasaran adalah pola pikir pemasaran yang akan digunakan untuk mencapai tujuan pemasarannya. Strategi pemasaran berisi strategi spesifik untuk pasar sasaran, penetapan posisi, bauran pemasaran, dan besarnya pengeluaran pemasaran [4].

Menurut Kotler bahwa promosi meliputi semua alat yang terdapat dalam bauran promosi yang peranan utamanya adalah komunikasi yang bersifat membujuk calon pembeli atau konsumen barang dan jasa [4].

Berdasarkan pengertian tentang strategi pemasaran di atas dapat disimpulkan bahwa sebuah strategi pemasaran merupakan salah satu upaya yang dapat dilakukan oleh sebuah sekolah dalam peningkatan penerimaan jumlah calon siswa baru di tahun - tahun berikutnya.

Data Mining

Secara sederhana, *data mining* dapat diartikan sebagai proses mengekstrak atau menggali *knowledge* yang ada pada sekumpulan data. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, seperti manajemen bisnis, pendidikan, kesehatan, dan sebagainya. Proses menggali informasi dalam *data mining* melibatkan integrasi

teknik dari berbagai disiplin ilmu, seperti teknologi *database* dan *data warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, *pattern recognition*, *neural network*, visualisasi data dan sebagainya [5].

Dalam rangka menemukan, menggali, atau menambang pengetahuan, terdapat enam fungsi alam data mining, yaitu: 1) Fungsi deskripsi (description), 2) Fungsi estimasi (estimation), 3) Fungsi prediksi (prediction), 4) Fungsi klasifikasi (classification), 5) Fungsi pengelompokan (classification), dan 6) Fungsi asosiasi (association) [6].

Mengacu kepada keenam fungsi data mining tersebut dapat dipilih menjadi : 1) Fungsi minor atau fungsi tambahan, yang meliputi ketiga fungsi yang pertama, yaitu deskripsi, Estimasi, dan prediksi; dan 2) Fungsi mayor atau fungsi utama, yang meliputi ketiga fungsi berikut, yaitu klasifikasi, pengelompokan, dan asosiasi [6].

Clustering

Madhu Yedha mendefinisikan *clustering* sebagai proses pengorganisasian objek data ke dalam *set* kelas yang saling berhubungan, yang disebut *cluster*. *Clustering* merupakan contoh dari klasifikasi tanpa arahan (*unsupervised*). Klasifikasi merujuk kepada prosedur yang menetapkan objek data set kelas. *Unsupervised* berarti bahwa pengelompokan tidak tergantung pada standar kelas dan pelatihan atau *training*. [7].

Menurut Deka, *Clustering* merupakan salah satu teknik *data mining* yang digunakan untuk mendapatkan kelompok-kelompok dari objek-objek yang mempunyai karakteristik yang umum di data yang cukup besar. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* atau grup sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode *unsupervised learning*. Menurut Oyelade, *clustering* dapat dibagi menjadi dua, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [8].

Algoritma K-Means

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokan data dalam bentuk satu atau lebih cluster / kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu cluster / kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan dengan cluster / kelompok yang lain sehingga data yang berada dalam satu cluster / kelompok memiliki tingkat variasi yang kecil [9].

Menurut Daniel dan Eko, Langkah-langkah algoritma *K-Means* adalah sebagai berikut :

- a. Pilih secara acak k buah data sebagai pusat *cluster*.
- b. Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*.

Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- c. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*.
- d. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat.
- e. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulangi sampai nilai *centroid* tidak berubah lagi [10].

Metode Penelitian

Analisa kebutuhan

Pada analisa kebutuhan sistem membahas beberapa kebutuhan dan atau persyaratan terkait dengan input, proses, dan output. Kebutuhan atau persyaratan ini diperoleh berdasarkan data siswa. Berdasarkan data tersebut diperoleh hasil analisa kebutuhan sistem berikut :

a. Kebutuhan input

Sistem membutuhkan data input yang berisi data siswa, antara lain : Nomor Induk Siswa (NIS), Nama Siswa, Tempat Tanggal Lahir, Jenis Kelamin, No. Hp, Alamat, Asal Sekolah, dan Program Studi.

b. Kebutuhan Proses

Proses yang digunakan untuk mengolah data input adalah teknik klustering dengan algoritma *K-Means*. Namun dalam penelitian ini hanya 30 data siswa yang di ambil sebagai sampel.

c. Kebutuhan Output

Output yang diharapkan adalah analisa dari hasil klustering perhitungan manual dengan implementasi *K-Means Clustering* sehingga dapat diketahui informasi / pola dari data input.

Perancangan / Desain

Proses perancangan / desain meliputi penyusunan data untuk memudahkan pada saat dilakukan pengklasteran. Secara umum penelitian mempertimbangkan dari segi proses perhitungan / komputasi dengan benar.

Pengujian / Testing

Proses pengujian meliputi ; data akademik siswa beserta data penerimaan siswa baru sebagai masukan (input) sehingga menghasilkan beberapa kluster dan analisisnya sesuai dengan rancangan.

Untuk tahap terakhir yaitu pemeliharaan / *Maintenance*, tidak dilakukan karena keterbatasan waktu.

Metode Penelitian

Data primer adalah data yang diperoleh dari hasil penelitian. Data primer dalam penelitian ini adalah data siswa kelas XII. Sedangkan untuk data sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Metode penelitian yang dilakukan adalah metode penelitian eksperimen menggunakan algoritma *K-Means Cluster* dengan tahapan penelitian seperti berikut :

Pengolahan Data Awal

Data didapatkan dari Tata Usaha SMK Terpadu Lampung Subang, dengan mengambil data berupa atribut Nama, Program Studi, Alamat, dan Nilai Semester.

No	NIS	Nama Siswa	Prodi	Sekolah Asal	Nilai Semester
1	17180001	Dandi Alawiyah	TKJ	Lampung	80
2	17180002	Dian Romansyah	TKJ	Parung	83
3	17180003	Geri Soluna	TKJ	Sadang	80
4	17180004	Haris	TKJ	Tanjungwangi	80
5	17180005	Hilman Maulana	TKJ	Parung	80
6	17180006	Indrik	TKJ	Caringin	85
7	17180007	Irvan Maulana	TKJ	Sadang	83

8	17180008	Japar Sidiq	TKJ	Lampang	83
9	17180009	Justian	TKJ	Parung	85
10	17180010	Nabila Luciana	TKJ	Parung	85
11	17180011	Riki	TKJ	Lampang	80
12	17180012	Rina Anggraeni	TKJ	Caringin	85
13	17180013	Said Hawa	TKJ	Lampang	80
14	17180014	Sri Cinta Noviyanti	TKJ	Sadang	85
15	17180015	Sugih Romadhan	TKJ	Parung	85
16	17180021	Ade Kurnia	TKJ	Tanjungwangi	83
17	17180022	Aisyah Awalia	TKJ	Sadang	85
18	17180023	Anita Pertiwi	TKJ	Lampang	85
19	17180024	Astri Nurlaela	AP	Sadang	80
20	17180025	Bayu Permana Mulya	AP	Parung	80
21	17180026	Cica	AP	Tanjungwangi	85
22	17180027	Cecep Diki Liana	AP	Lampang	80
23	17180028	Dini Agustiani	AP	Caringin	85
24	17180098	Ejang Hendarsah	AP	Tanjungwangi	83
25	17180029	Eka Yulianti	AP	Lampang	83
26	17180030	Eva Nurlaela	AP	Tanjungwangi	85
27	17180031	Fina Nurhasanah	AP	Parung	85
28	17180032	Hardani Muhayad	AP	Sadang	80
29	17180100	Mochammad Rizky R.	AP	Caringin	85
30	17180033	Nova Mardiana	AP	Lampang	83

Tabel 1 : Data Siswa SMK Terpadu Lampang Subang
(Sumber : Tata Usaha SMK Terpadu Lampang Subang)

Transformasi Data

Pada tahap ini dilakukan proses perubahan data, tujuannya adalah agar data di atas dapat diolah dengan menggunakan metode *K-Means Clustering*. Data yang berjenis data nominal seperti alamat dan program studi harus diinisialisasikan terlebih dahulu dalam bentuk angka / numerikal. Untuk melakukan inisialisasi alamat dapat dilakukan dengan :

1. Pada alamat siswa terlebih dahulu dilakukan pembagian wilayah – wilayah menjadi beberapa bagian.
2. Kemudian wilayah tersebut di urutkan dari yang terbesar berdasarkan frekuensi mahasiswa yang berasal dari alamat tersebut.
3. Setelah itu wilayah yang memiliki frekuensi terbesar diberi inisial dengan angka 1, dan wilayah yang memiliki frekuensi terbesar kedua diberi inisial dengan angka 2, begitu seterusnya hingga wilayah dengan frekuensi paling sedikit. Hasil dari inisialisasi alamat dapat dilihat pada tabel berikut.

Wilayah Alamat	Inisial
Lampang	1
Parung	2
Sadang	3
Tanjungwangi	4
Caringin	5

Tabel 2 : Inisialisasi Data Alamat Siswa SMK Terpadu Lampang

Selain kota asal, program studi juga termasuk ke dalam jenis data nominal. Sehingga perlu di inisiasikan ke dalam bentuk angka. Seperti pada alamat, pada program studi juga diberikan inisiasi berdasarkan frekuensi mahasiswa.

Prodi	Inisial
TKJ	1
AP	2

Tabel 3 : Inisialisasi Data Prodi Siswa SMK Terpadu Lampung

Pengolahan Data

Setelah semua data mahasiswa ditransformasikan ke dalam bentuk angka, maka data – data tersebut telah dapat dikelompokkan dengan menggunakan *K-Means Clustering*.

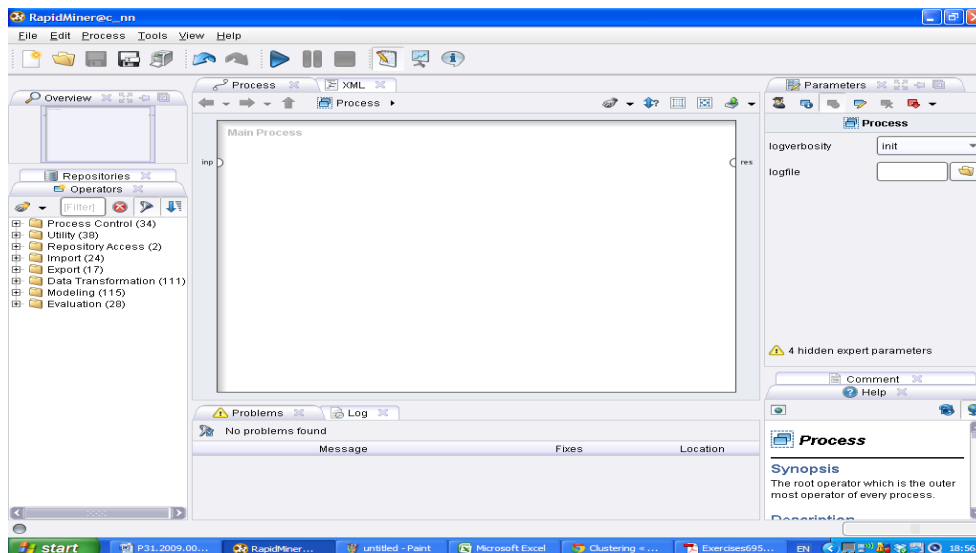
Hasil dan Pembahasan

Data Uji

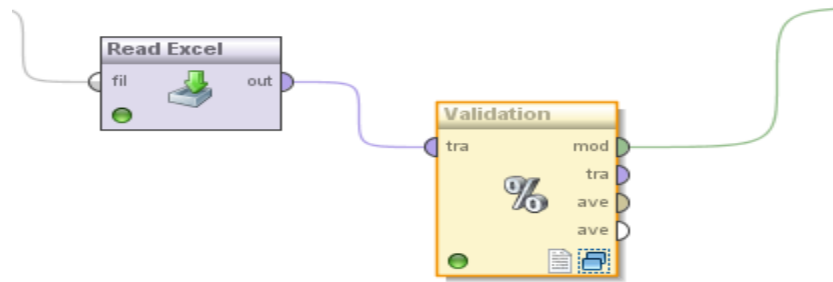
Data yang digunakan dalam penelitian ini merupakan data siswa kelas XII SMK Terpadu Lampung Subang angkatan 2017/2018. Data siswa yang digunakan memiliki tiga atribut yang digunakan untuk proses pengklasteran, diantaranya adalah Program Studi, Alamat, dan Nilai Semester.

Pengujian Model

Pada penelitian ini *rules* atau model yang terbentuk diterapkan pada program yang diperoleh dengan bantuan *tools RapidMiner*. Dengan menggunakan pemodelan *K-Means Clustering*, data di inisialisasi dengan jumlah cluster sebanyak 2 buah, maka didapat hasil dengan *cluster* yang terbentuk dari data sebanyak 30 record dengan 3 atribut dimana semua atribut bertipe kategorikal. Berikut gambar *tools RapidMiner* untuk menghitung pengklasteran :



Gambar 1 : Tampilan *Tools Rapid Minner*



Gambar 2 : *Split validation* dengan *K-Means Clustering*

Attribute	Cluster_1	Cluster_2	Cluster_3
Prodi	1,75	1,72	1,75
Alamat	10,5	23,5	23,5
Nilai Semester	3,36	3,5	3,47

Tabel 4 : Hasil perhitungan manual antara jarak cluster

Hasil *Clustering*

Proses melakukan export data ke excel mempermudah pada saat melakukan pengolahan data dan evaluasi terhadap hasil dari cluster yang terbentuk saat melakukan implementasi pada tools *rapidminer*.

Berdasarkan hasil pengelompokkan data menggunakan metode *K-Means Clustering*, di dapatkan hasil dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*.

Pembahasan

Dari hasil *cluster 1*, terlihat bahwa karakteristik siswa pada cluster 1 di dominasi siswa yang berasal dari program studi TKJ. Sedangkan berdasarkan alamat di dominasi oleh siswa yang beralamat di Lampung. Sehingga dapat disimpulkan bahwa rata – rata siswa pada *cluster 1* yang berasal dari wilayah alamat Lampung mengambil program studi TKJ.

Dari hasil *cluster 2*, terlihat bahwa karakteristik siswa pada cluster 2 di dominasi siswa yang berasal dari program studi AP. Sedangkan berdasarkan alamat di dominasi oleh siswa yang beralamat di Tanjungwangi. Sehingga dapat disimpulkan bahwa rata – rata siswa pada *cluster 1* yang berasal dari wilayah alamat Tanjungwangi mengambil program studi AP.

Sedangkan Dari hasil *cluster 3*, terlihat bahwa karakteristik siswa pada cluster 3 di dominasi siswa yang berasal dari program studi TKJ. Sedangkan berdasarkan alamat di dominasi oleh siswa yang beralamat di Parung. Sehingga dapat disimpulkan bahwa rata – rata siswa pada *cluster 3* yang berasal dari wilayah alamat Parung mengambil program studi TKJ.

Kesimpulan dan Saran

Kesimpulan

Berdasarkan hasil penelitian pada pengklasteran maka dapat diambil beberapa kesimpulan sebagai berikut :

1. *K-Means Clustering* dapat mengklaster data siswa untuk strategi terbaik marketing sekolah. Dengan cara melakukan promosi dengan mengirim tim marketing yang sesuai dengan program studi yang banyak diminati.
2. Melakukan promosi pada wilayah – wilayah yang ada di sekitar sekolah sesuai dengan alamat siswa.

3. Pengujian yang dilakukan ditemukan pusat *cluster* : *Cluster* 1 = 1,75; 10,75; 3,36. *Cluster* 2 = 1,72; 1,72; 3,50, dan *Cluster* 3 = 1,75; 23,50; 3,47.
4. Hasil *Cluster* berdasarkan program studi adalah bahwa program studi TKJ menduduki posisi pertama yang paling banyak diminati.
5. Promosi akan diperbanyak pada hasil *Cluster* terendah.

Saran

1. Dengan adanya penerapan algoritma *K-Means Cluster* diharapkan mampu memberikan solusi dan dapat membantu SMK Terpadu Lampung Subang dalam melaksanakan marketing sekolah.
2. Pengelompokkan data sebaiknya dilakukan secara rutin setiap tahunnya.
3. Penelitian dapat dikembangkan dalam proses penentuan strategi promosi penerimaan siswa baru.

DAFTAR PUSTAKA

- Ediyanto, dkk, Pengklasifikasian Karakteristik dengan Metode *K-Means Cluster* Analisis Sistem. Buletin ilmiah Mat. Stat dan Terapannya (Bimaster), Volume 02 No 2. 2013
- Johan Oscar Ong, Implementasi Algoritma *K-Means Clustering* untuk Menentukan Strategi Marketing President University. Jurnal ilmiah teknik industri Vol. 12 No 1, 2013.
- T. Gomasthit, *Cloud Coverage Identification Using Satellite Data and K-Means Clustering Algorithm*. *Journal of Global research in computer Science*, volume 4 No. 7, 2013.
- [4] Kotler Philip, manajemen pemasaran (analisis, perencanaan, pengendalian)terjemahan drs. jaka. wasana msm. Jakarta : Erlangga, 1996.
- [5] Agus Nur Khomarudin. Teknik *Data Mining* : Algoritma *K-Means Clustering*. IlmuKomputer.Com 2003-2016.
- [6] Larose, D.T. *Introduction to data mining* (pp. 1 - 26). John Wiley & Sons, Inc. 2005.
- [7] Yedla, M., Pathakota, S. R. and Srinivasa, T. M. “*Enhancing K-means Clustering Algorithm with Improved Initial Center.*” *International Journal of Computer Science and Information Technologies*. 1. 121. 2010.
- [8] Deka Dwinavinta Candra Nugraha, Zumrotun Naimah, Makhfuzi Fahmi dan Novi Setiani.. “Klasterisasi Judul Buku dengan Menggunakan Metode K-Means.” *Seminar Nasional Aplikasi Teknologi Informasi*. ISSN: 1907-5022. G-2. 2014
- [9] Giyanto, Heribertus. *Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva*. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada. 2008
- [10] Daniel Riano Kaparang dan Eko Sedyono. “Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis Algoritma K-means di Wilayah Kabupaten Boyolali.” *JdC*. 2. 20, 2013.
- [11] Siti N, *Sistem Pendukung Keputusan Untuk Menentukan Strategi Marketing Kampus Menggunakan Metode K-Means Clustering*. Subang, 2017.