



Analisis Nilai Rapor Siswa Kelas X Jurusan Multimedia Terhadap Minat pada Pelajaran Produktif Di Kelas XII untuk Menentukan Kompetensi Siswa dengan Metode *Clustering Algoritma K-Means* (Studi Kasus Di SMKN 4 Padang)

Netri Elizawati¹ dan Lido Sabda Lesmana²

¹SMK Negeri 4 Padang, email: netrie78@gmail.com

²Universitas Putera Batam, email: lidosabdalesmana11603@gmail.com

Abstrak

Salah satu teknik yang bisa digunakan dalam Data mining adalah Clustering. Clustering adalah teknik utama yang digunakan untuk pengelompokan data numerik dan gambar dalam aplikasi data mining dan pengolahan gambar. Data Mining merupakan proses ekstraksi data menjadi informasi, yang dapat membantu perusahaan dalam mendapatkan pola dari data yang tersimpan didalam basis data. data yang dulunya dianggap tidak berguna bisa dimanfaatkan untuk pengambilan keputusan baik untuk kepentingan bisnis maupun pendidikan. Dalam dunia pendidikan data mining perlu diterapkan untuk membantu proses pengolahan data salah satunya dalam hal pengelompokan data siswa. Data yang biasanya hanya dijadikan sebagai arsip disekolah dapat dijadikan sebagai pedoman oleh para guru dalam hal penentuan minat siswa. Hal tersebut juga diterapkan pada SMKN 4 Padang yang akan mengelompokkan sata siswa berdasarkan minat siswa terhadap mata pelajaran produktif Multimedia sehingga dapat digunakan untuk menentukan kompetensi siswa dan dapat juga digunakan dalam pengambilan keputusan untuk menentukan penguatan materi yang akan dipelajari di kelas XII. Data yang digunakan adalah data nilai rapor siswa kelas X Multimedia tahun masuk 2013 semester I dan II serta credit title tugas mata pelajaran produktif siswa saat mereka kelas XII. Penelitian ini akan mencari kompetensi siswa berdasarkan minat pada mata pelajaran produktif. Metode yang digunakan untuk mengekstrasi data tersebut adalah metode clustering dengan menggunakan algoritma K-Means. Algoritma K-Means merupakan salah satu metode Clustering yaitu model centroid. Maksudnya menggunakan centroid untuk membuat cluster. Centroid adalah "titik tengah" suatu cluster. Suatu objek data termasuk dalam suatu cluster jika memiliki jarak terpendek dengan centroid data tersebut

Kata kunci: Data Mining, Clustering, Algoritma K-means, SMKN 4 Padang.

Abstract

One of the techniques that can be used in Data mining is Clustering. Clustering is the main technique used for grouping numerical data and images in data mining applications and image processing. Data Mining is the process of data extraction into information, which can help companies in getting the pattern of data stored in data base. data previously considered useless

can be used for decision making both for business and education. In the world of education data mining needs to be applied to help the data processing one of them in terms of grouping student data. Data that usually only used as a school archive can be used as a guide by teachers in terms of determination of student interest. It is also applied to SMKN 4 Padang that will classify sata students based on students' interest on Multimedia productive subjects so that it can be used to determine the competence of students and can also be used in decision making to determine the strengthening of material to be studied in class XII. The data used is the data value of student report cards class X Multimedia year entering 2013 semester I and II and the credit title of students 'productive subjects' tasks as they class XII. This study will look for student competencies based on interest in productive subjects. The method used to extract the data is clustering method by using K-Means algorithm. K-Means algorithm is one of Clustering method that is centroid model. It means using a centroid to create a cluster. Centroid is the "midpoint" of a cluster. A data object is included in a cluster if it has the shortest distance with the centroid of the data.

Keywords: Data Mining, Clustering, Algoritma K-means, SMKN 4 Padang.

1. Pendahuluan

Ilmu pengetahuan dan teknologi sekarang ini, berkembang pula konsep Data Mining untuk mengolah data yang dulunya dianggap sampah, sehingga data-data yang tersimpan dan terakumulasi dalam jumlah yang banyak bisa digali dan dimanfaatkan untuk pengambilan keputusan baik untuk kepentingan bisnis maupun pendidikan. Pendapat lain tentang data mining menyebutkan bahwa Data Mining merupakan proses ekstraksi data menjadi informasi yang sebelumnya belum tersampaikan, dengan teknik yang tepat proses data mining akan memberikan hasil yang optimal [1].

Dalam metode *Clustering* banyak teknik yang dapat digunakan, salah satunya adalah penggunaan Algoritma K-Means. Menurut tinjauan sebuah jurnal, penggunaan algoritma K-Means dapat membantu dalam mengelompokkan data, informasi yang ditampilkan berupa nilai centroid dari tiap-tiap cluster. Untuk mencegah terjadinya perpindahan pelanggan, maka perlu diketahui kelompok pelanggan potensial, sehingga perusahaan dapat memberikan pelayanan prima [2].

Algoritma K-Means ini sudah banyak diterapkan dalam penelitian di berbagai bidang ilmu, seperti "Implementasi Algoritma K-Means untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus :SMP Negeri 101 Jakarta)" [3]. kesimpulan dari penelitian ini adalah algoritma K-Means dapat diimplementasikan untuk membantu pengelompokkan kemampuan siswa terhadap mata pelajaran Ujian Nasional. Sedangkan pada penelitian dengan judul "Penerapan Metode K-Means untuk *Clustering* Produk *Online Shop* dalam Penentuan Stok Barang" [4] disimpulkan bahwa berdasarkan pengolahan data dapat dikelompokkan produk yang diminati dan kurang diminati. Untuk penelitian lain yang berjudul "Penerapan Data Mining Pemakaian Air Pelanggan untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma K-Means" [1] disimpulkan bahwa metode K-Means *Clustering* dapat digunakan untuk mengelompokkan potensi pendapatan dan pemakaian air.

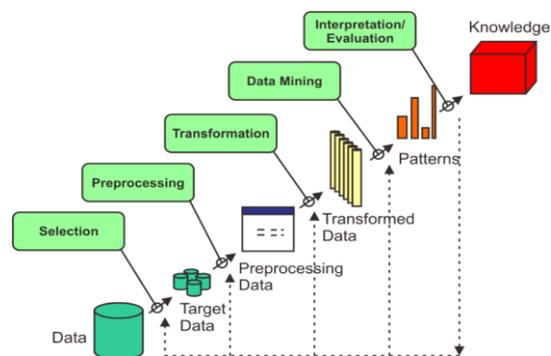
Pola yang dihasilkan dari algoritma ini adalah sebuah pengelompokkan data siswa berdasarkan minat terhadap mata pelajaran produktif Multimedia, sehingga dapat digunakan untuk menentukan kompetensi siswa, dan dapat juga digunakan dalam

pengambilan keputusan untuk menentukan penguatan materi yang akan dipelajari di kelas XII. *Output* dari penelitian ini adalah siswa yang berminat sebagai *movie maker* atau siswa yang berminat sebagai artist.

2. Landasan Teori

2.1 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) biasanya disamakan dengan Data Mining, berdasarkan tujuannya dalam mencari pengetahuan yang berguna dalam tumpukan data yang besar. Data Mining sebenarnya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik [5]. Pendapat yang senada adalah Data Mining disebut juga *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data [6].



Gambar 1. Tahapan Proses KDD dalam Database

2.2 Data Mining

Pengertian Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database [2], pendapat lainnya adalah Data Mining merupakan proses ekstraksi data menjadi informasi yang sebelumnya belum tersampaikan [1].

Data Mining merupakan proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya [5]. Pendapat lain tentang Data mining adalah *Data mining is the process of combining the different data source and derives the new pattern from that data collection* [7].

2.2.1 Proses Data Mining

Proses dalam Data mining adalah sebagai berikut : [1]

1. Pembersihan data (*data cleaning*), yaitu proses menghapus data pengganggu (noise) yang dikatakan tidak konsisten atau tidak diperlukan.
2. Integrasi data (*data integration*), yaitu menggabungkan berbagai sumber data.
3. Pemilihan data (*data selection*), yaitu memilih data yang dipilih sesuai kebutuhan analisis.
4. Transformasi data (*data transformation*), yaitu proses transformasi data ke dalam format untuk diproses dan siap ditambang.

5. Penggalian data (*data mining*), yaitu menerapkan metode kecerdasan untuk ekstraksi pola.
6. Evaluasi pola (*pattern evaluation*), yaitu mengidentifikasi pola-pola yang menarik yang merepresentasikan pengetahuan.
7. Penyajian pola (*knowledge presentation*), yaitu teknik untuk memvisualisasikan pola pengetahuan ke pengguna.

2.3 Clustering

Clustering yaitu proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelas atau kelompoknya, dengan kata lain *Clustering* proses untuk mengetahui kelas-kelas atau taksonomi atau analisis topologi dari data-data yang ada.

“Konsep dasar dari *Clustering* adalah mengelompokkan sejumlah objek ke dalam *Cluster* dimana *Cluster* yang baik adalah *Cluster* yang memiliki tingkat kesamaan yang tinggi dengan objek *Cluster* yang lainnya [1]. Terdapat berbagai macam algoritma *Clustering*, namun dalam penggunaannya tergantung pada tipe data yang akan dikelompokkan. Klasifikasi algoritma *Clustering* dapat dibagi menjadi menjadi 5 kategori, yaitu: [1]

1. *Partitioning methods*

Pengelompokkan objek dimana tiap objek dimiliki oleh 1 cluster. Algoritma partisi dapat meminimalkan mean kuadrat jarak dari setiap titik data ke pusat terdekatnya. Yang termasuk ke dalam metode ini adalah algoritma k-means, k-medoid atau PAM, CLARA, dan CLARANS. PAM

2. *Hierarchical methods*

Pengelompokkan objek dapat dilakukan dengan 2 cara, agglomerative yang dimulai dengan menggabungkan beberapa cluster hingga menjadi satu, atau divisive yang dimulai dengan cluster yang sama kemudian dipecah menjadi beberapa cluster yang lebih kecil. Yang termasuk ke dalam metode ini adalah algoritma CURE, BIRCH, dan Chameleon.

3. *Density-based methods*

Pengelompokkan objek berdasarkan tingkat kerapatan objek atau densitas. Yang termasuk dalam metode ini adalah algoritma DBSCAN, DENCLUE, dan OPTICS.

4. *Grid-based methods*

Pengelompokkan objek dengan menggunakan struktur data grid multiresolusi yang mampu menangani data berdimensi tinggi. Yang termasuk dalam metode ini adalah algoritma CLIQUE, WaveCluster, dan STING.

5. *Model-based methods*

Pengelompokkan objek dengan memodelkan tiap cluster, dan mencoba mengoptimasikan kesesuaian data dengan model matematika. Yang termasuk ke dalam metode ini adalah algoritma COBWEB.

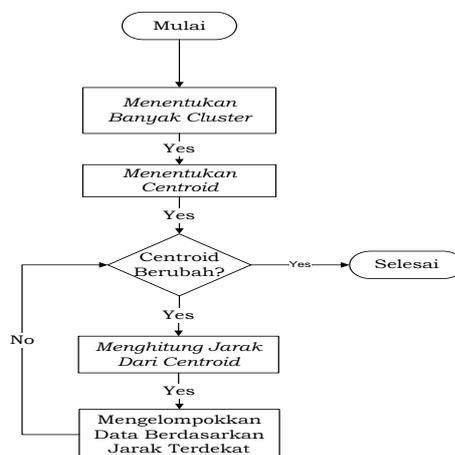
2.4 K-Means

K-Means *Clustering* diperkenalkan oleh Maquen tahun 1967. K-Means *Clustering* adalah satu bentuk yang paling sederhana dari algoritma pembelajaran tanpa pengawasan yang digunakan untuk memecahkan masalah berdasarkan cluster [7].

K-Means merupakan salah satu metode pengelompokan data *nonhierarki* (sekatan) yang berusaha mempatisi data yang ada ke dalam bentuk dua atau lebih kelompok [2].

Algoritma K-Means merupakan salah satu metode *Clustering* yaitu model centroid. Maksudnya menggunakan centroid untuk membuat cluster. Centroid adalah “titik tengah” suatu cluster. Suatu objek data termasuk dalam suatu cluster jika memiliki jarak terpendek dengan centroid data tersebut. Ada dua tahapan dalam algoritma K-Means, yaitu : [7]

1. Perbaiki nilai K di muka dan memilih pusat K dicara acak.
2. Menetapkan objek data ke pusat terdekat



Gambar 2. Flowchart dari algoritma K-Means

Tahapan dalam algoritma K-Means adalah :

1. Menentukan banyaknya cluster (k). Jumlah cluster harus lebih kecil dari banyaknya data ($k < n$).
2. Tentukan centroid setiap cluster. Untuk menentukan centroid awal, banyak metode yang dapat digunakan, antara lain dengan mengambil dari data sumber secara acak atau random.
3. Hitung jarak data dengan centroid. Rumus untuk menghitung jarak menggunakan rumus “Euclidean”.

Dalam Penelitian tentang menghitung jarak dengan rumus, Rajalakshmi [7]

$$d(x,y)=||x-y||= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1,2,3,\dots,n \quad (1)$$

Di mana :

x_i : objek x ke- i

y_i : daya y ke- i

n : banyaknya objek

4. Kelompokkan data sesuai dengan clusternya, yaitu cluster terpendek.
5. Ulangi langkah 2 sampai ditemukan kriteria yang diinginkan.

2.5 Perbandingan Penelitian

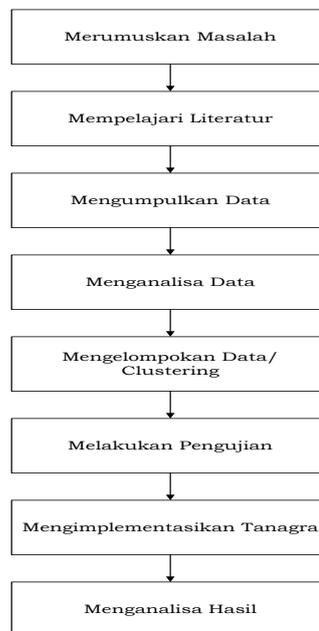
Tabel 1. Perbandingan Penelitian tentang algoritma K-Means

NO	Nama Peneliti	Judul Penelitian	Metode	Hasil
1	Angga Ginanjar dan Mabrur Riani Lubis (2012)	Penerapan data mining untuk memprediksi kriteria nasabah kredit	<i>Teknik data mining dengan metoda klasifikasi menggunakan Decision Tree Algoritma C 4.5</i>	Aplikasi yang dibangun dapat membantu Bagian Dana pada Bank XY dalam menganalisis data nasabah untuk menentukan target pemasaran kredit sehingga diharapkan biaya operasional marketing perbankan dapat ditekan seminimal mungkin.
2	Fenty Eka M. Agustin, et al, 2015	Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus :SMP Negeri 101 Jakarta)	<i>K-Means Clustering</i>	algoritma K-Means dapat diimplementasikan untuk membantu pengelompokkan kemampuan siswa terhadap mata pelajaran Ujian Nasional
3	Elly Muningsih dan Sri Kriswati, 2015	Penerapan Metode K-Means untuk Produk Online Shop dalam Penentuan Stok Barang	<i>K-Means Clustering</i>	Berdasarkan pengolahan data dapat dikelompokkan produk yang diminati dan kurang diminati
4	Gunawan Abdillah, et al, 2016	Penerapan Data Mining Pemakaian Air Pelanggan untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma K-Means	<i>K-Means Clustering</i>	Metode K-Means <i>Clustering</i> dapat digunakan untuk mengelompokkan potensi pendapatan dan pemakaian air.

5	Fina Nasari dan Surya Darma, 2015	Penerapan K-Means Clusteing pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)	Algoritma Means	K-	Kesimpulan dari penelitian ini adalah bahwa jika asal sekolah adalah SMA atau SMP maka rata-rata jurusan yang diambil adalah Sistem Informasi dan jika asal Sekolahnya adalah SMK rata-rata jurusan yang diambil adalah Teknik Informatika
---	-----------------------------------	---	-----------------	----	--

Kerangka Kerja

Dalam melakukan penelitian, perlu dibuat sebuah kerangka kerja (*Framework*) agar pekerjaan yang dilakukan dapat lebih terstruktur dan tertata dengan baik. Adapun kerangka kerja dalam penelitian ini dapat dilihat pada gambar 3:



Gambar 3. Kerangka Kerja Penelitian

Dari gambar 3 dapat dijabarkan kerangka kerja yang akan dilakukan pada penelitian ini adalah sebagai berikut :

1. Merumuskan Masalah

Hal yang paling mendasar dalam sebuah penelitian adalah bagaimana melakukan analisa terhadap sebuah kasus.

2. Mempelajari Literatur

Literatur didapatkan dari jurnal ilmiah baik jurnal nasional maupun jurnal internasional, dari buku, *paper*, artikel-artikel, informasi dari internet dan sumber-sumber lainnya yang relevan dengan metode yang digunakan, yaitu

berkaitan dengan KDD, *Data Mining*, *Clustering*, Algoritma K-Means dan sebagainya.

3. Mengumpulkan Data

Dalam tahapan ini yang akan dilakukan adalah pengumpulan terhadap data dan informasi mengenai masalah dalam topik penelitian. Untuk memperoleh data dan informasi ini metode yang digunakan adalah sebagai berikut :

a. Metode Wawancara

Dalam proses wawancara ini penulis melakukan tanya jawab dengan pihak-pihak yang terkait dengan data yang penulis butuhkan, antarlain walikelas, admin sisfo, ataupun pimpinan dalam hal ini kepala sekolah SMKN 4 Padang sebagai pengambil kebijakan mengenai penginputan data, kelengkapan data, ketersediaan data, penyimpanan data, pemanfaatan data untuk pengambilan keputusan dan sebagainya.

b. Metode Observasi

Observasi dilakukan langsung di SMKN 4 Padang, hal ini dilakukan pada saat wawancara dan pengambilan data. Hal ini bertujuan untuk menganalisa permasalahan dan apakah topik penelitian ini sesuai dengan permasalahan yang terjadi di sekolah tersebut.

c. Studi Pustaka

Pengumpulan data dan penelusuran informasi dilakukan dengan cara membaca dan mempelajari buku-buku dan jurnal-jurnal penelitian yang berkaitan dan menunjang, baik dalam penganalisaan data dan informasi, maupun pemecahan masalah secara keseluruhan.

4. Menganalisa Data

Setelah data mentah didapatkan perlu dilakukan analisa dan penyeleksian terhadap data tersebut sebelum dilakukan pengolahan lebih lanjut. Pemilihan terhadap data-data mana yang tidak diperlukan dan harus dibuang dan mana data-data yang dibutuhkan. Jika dalam penelitian ada data yang perlu dikelompokkan, maka pengelompokkan data tersebut itu dilakukan pada tahap ini, sehingga data yang ada siap untuk diproses dan diolah sesuai dengan yang dibutuhkan nantinya.

5. Pengelompokkan Data / *Clustering*

Data yang telah dibersihkan dan diseleksi akan menghasilkan data yang baik untuk proses selanjutnya. Kemudian, dari data tersebut dibuat kelompoknya sesuai dengan teori yang telah dipelajari.

6. Melakukan Pengujian

Sebelum dilakukan pengujian menggunakan *software* Tanagra, perlu dilakukan terlebih dahulu pengujian secara manual, sesuai dengan teori dan rumus-rumusny. Pengujian manual dilakukan terhadap data yang dijadikan sampel, yang kira-kira bisa mewakili data secara keseluruhan. Setelah dilakukan pengujian manual, kemudian dilakukan pengujian dengan menggunakan *software*. Keputusan yang dihasilkan dari pengolahan data menggunakan *software* nantinya akan diuji aturannya, apakah aturan yang dihasilkan sudah sesuai dengan apa yang diharapkan atau sebaliknya, dan apakah aturan-aturan tersebut dapat digunakan untuk kepentingan pendidikan dan penelitian lainnya.

7. Mengimplementasikan Tanagra

Pada tahap ini variabel yang sudah diperoleh dan data yang sudah dibersihkan diproses dengan menggunakan bantuan salah satu *software Data Mining* yaitu Tanagra. Tanagra merupakan salah satu *software Data Mining* yang populer dan cukup banyak digunakan, baik untuk kepentingan bisnis oleh perusahaan ataupun untuk pendidikan dan penelitian. Pengolahan dan penggunaan *software* ini dilakukan pada sebuah laptop dengan spesifikasi sebagai berikut :

- a. *Processor* : Intel Core i3
- b. *Memory* : DDR3 3 Gb
- c. *VGA* : Intel(R) HD Graphics 3000
- d. *Harddisk* : 500Gb

Sedangkan perangkat lunak yang digunakan adalah sebagai berikut :

- a. *Sistem Operasi* : *Microsoft Windows 8, 32Bit*
- b. *Aplikasi Office* : *Microsoft Office 2007*
- c. *Software Data Mining* : *Tanagra*

8. Menganalisa Hasil

Data yang telah diolah dengan menggunakan *software* Tanagra akan menghasilkan *Clustering* kompetensi siswa, sehingga hasil tersebut nantinya dapat dimanfaatkan untuk keperluan ataupun untuk ilmu pengetahuan.

3. Proses Clustering dengan algoritma K-Means

K-Means termasuk dalam metode data mining yang memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain. *K-Means* memisahkan data ke *k* daerah bagian yang terpisah, dimana *k* adalah bilangan integer positif. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengklasifikasi data besar.

Berikut adalah langkah-langkah Algoritma *K-Means* :

1. Menentukan jumlah *cluster*.
2. Menentukan pusat *cluster* awal.

Dalam menentukan pusat *cluster* awal dilakukan pembangkitan bilangan *random* yang merepresentasikan urutan data input. Pusat *cluster* awal didapatkan dari data sendiri bukan dengan menentukan titik baru.

3. Menghitung jarak dengan pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian Distance*, algoritma perhitungan jarak data dengan pusat *cluster* adalah sebagai berikut :

Ambil nilai data dan nilai pusat cluster.

Hitung *Euclidian Distance* data dengan tiap pusat *cluster* dengan rumus, Rajalakshmi

[7] :

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1,2,3,\dots,n \quad (2)$$

Di mana :

- x_i : objek x ke- i
- y_i : daya y ke- i
- n : banyaknya objek

4. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat.

Algoritma pengelompokan data :

- a. Ambil nilai jarak tiap pusat *cluster* dengan data
- b. Cari nilai jarak terkecil
- c. Kelompokan data dengan pusat *cluster* yang memiliki jarak terkecil.

5. Penentuan pusat *cluster* baru

Untuk mendapatkan pusat *cluster* baru bisa dihitung dari rata-rata nilai anggota *cluster* dan pusat *cluster*. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).

Algoritma penentuan pusat *cluster*, Rajalakshmi [7] :

- a. Cari jumlah anggota tiap *cluster*
- b. Hitung pusat baru dengan rumus

$$\text{Pusat } cluster \text{ baru} = \frac{\sum_{i=1}^n x_i}{n}; i = 1,2,3,\dots,n \quad (3)$$

Dimana :

- x_i : objek ke- i (anggota *cluster*)
- n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*

4. Implementasi

Tabel 2. Data Rata-Rata Nilai Semester I dan II

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	
1	ADEK NOFRINANDES	2,67	2,67	3,17	2,67	2,84	2,84	2,67	3,17	2,84	2,67	3,17	2,67	2,84	2,84	2,67	2,84	3,17
2	ADITYA ALHAMDY SYAHPUTRA	3,17	2,84	3,50	2,67	3,17	3,17	2,67	3,50	3,17	2,67	3,67	2,67	3,34	3,50	3,00	2,84	3,50
3	AL KAUTSAR	2,67	2,67	3,00	2,67	2,67	2,67	2,67	3,00	2,84	2,67	2,67	2,67	3,00	2,67	2,67	3,00	2,67
4	ANDREAN WIRAMUKTI F.	3,00	3,00	3,33	2,67	2,84	2,67	2,67	3,00	2,84	2,67	3,34	2,84	3,17	2,84	3,00	2,84	3,67
5	BENNY NOVRI RAHMAN	3,17	2,84	3,33	3,00	2,67	3,17	2,84	2,84	2,84	2,67	3,00	2,67	3,17	3,67	3,00	3,17	3,50
6	BENNY SYAHPUTRA	2,67	2,84	3,00	2,67	2,67	2,84	2,84	3,00	2,67	3,00	2,67	2,84	2,84	2,67	2,84	3,50	
7	CINDY AUDRY MONICA	3,17	2,84	3,50	2,67	3,00	3,00	2,84	3,17	3,00	3,34	3,50	3,00	3,17	2,84	3,33	3,34	3,84
8	DICKY PRATAMA JUNAIDI	3,17	2,84	3,66	3,00	3,34	3,17	2,67	3,33	3,17	3,00	3,67	3,33	3,17	3,17	3,33	3,17	3,67
9	FADHILAH TAMAMI	2,92	3,15	3,36	3,46	3,03	3,01	3,30	3,20	3,14	3,00	3,09	3,17	3,10	3,17	3,17	3,17	3,17
10	FAJRYSAN	3,00	2,84	3,17	2,67	2,84	3,17	2,67	3,50	3,00	2,67	3,17	2,67	3,00	2,67	3,00	3,00	3,50
11	FAUZI RIJAL	3,00	2,84	3,17	2,67	2,84	3,33	2,67	3,50	3,00	2,67	3,50	2,67	3,17	3,34	3,00	3,33	3,34
12	FENI SUNDARI PUTRI	2,67	3,17	3,50	2,67	3,00	2,67	2,67	3,33	2,67	2,67	3,17	2,84	3,00	3,00	3,00	3,17	3,33
13	GUSRINALDI	3,17	2,84	3,00	2,84	3,00	3,00	2,67	2,84	2,84	2,67	3,34	2,84	3,17	3,17	3,33	3,50	3,50
14	HADRILA PUTRI ASWARA	3,00	3,17	3,00	2,67	3,00	2,67	3,17	3,17	2,67	3,34	2,67	2,67	3,00	3,00	3,00	3,00	3,17
15	ILHAM ADE SAPUTRA	2,67	2,67	3,00	2,67	2,67	2,84	2,67	3,00	2,84	2,67	3,17	2,67	2,84	2,67	2,67	3,00	3,34
16	IMAM QADRI	2,67	2,67	3,00	2,84	2,84	2,84	2,67	3,34	2,67	2,84	3,50	2,67	3,00	2,84	2,67	3,34	3,50
17	INDAH NURUL FADIAH	2,67	3,17	3,00	2,67	3,00	2,67	2,67	3,33	2,84	2,67	3,17	2,67	2,84	2,67	2,67	2,84	3,33
18	MUHAMMAD IRSYAD	3,34	2,84	3,00	2,67	3,00	2,67	2,67	3,17	3,00	2,67	3,50	2,67	3,00	3,00	2,67	3,00	3,67
19	MUHAMMAD ILHAM CHATAMY	3,00	3,00	3,34	2,67	3,00	2,67	2,84	3,00	3,00	2,67	3,34	2,67	3,00	3,00	2,67	3,00	3,84
20	MUHAMMAD RIVA'I	3,00	2,84	3,00	2,67	3,00	2,84	3,00	3,33	3,17	2,67	3,50	2,67	3,00	3,34	3,00	3,17	3,50
21	MUTIARA VITONALIKA	2,84	3,00	3,33	2,67	3,00	3,67	2,84	2,84	3,34	2,67	3,00	2,67	3,17	3,00	3,33	3,00	3,33
22	NADIYA UTARI	3,00	3,17	3,33	2,67	3,00	3,50	3,00	3,00	3,17	2,84	3,34	2,84	3,00	2,84	3,00	3,00	3,50
23	PITRI MULIANY IHSAN	2,67	2,84	3,34	2,67	3,00	3,17	2,84	2,84	2,84	2,67	3,34	2,84	3,00	3,00	3,00	2,84	3,50
24	RADIN SEPTIAWAN AFNIM	3,34	3,67	3,67	3,00	3,50	3,50	3,34	3,50	3,50	2,84	3,67	3,34	3,33	3,34	3,33	3,50	4,00
25	RASYID EKA PUTRA	3,17	2,84	3,00	2,67	2,84	3,00	3,34	2,84	3,17	2,67	3,34	3,00	3,17	3,17	2,67	2,84	3,34
26	SRI MELDA FADHILLA	3,17	2,84	3,67	3,00	3,17	3,67	2,84	3,50	3,50	2,84	3,50	3,50	3,17	3,34	3,33	3,33	3,33
27	SUHAILA	3,17	3,67	3,17	3,17	3,00	2,67	2,84	3,50	3,17	2,67	3,50	3,17	3,17	2,84	3,00	2,84	3,34
28	UKHTI RAHMANADA ZIKRA	2,67	2,67	3,00	2,67	3,00	3,00	2,67	3,17	2,84	2,67	3,34	2,67	3,00	2,84	3,00	3,00	3,33

Tabel 2. diatas merupakan tabel rata-rata nilai rapor siswa kelas X Multimedia SMKN 4 Padang tahun masuk 2013. Data inilah yang akan dianalisa menggunakan algoritma K-Means. Percobaan dilakukan dengan menggunakan parameter-parameter berikut :

- Jumlah cluster : 2
- Jumlah data : 28
- Jumlah atribut : 17

Nama Atribut : Agama (X1), PPKN (X2), Bahasa Indonesia (X3), Matematika(X4), Sejarah(X5), Bahasa Inggris(X6), Seni Budaya(X7), PJOK(X8), PKWU(X9), Fisika (X10), Pemograman Dasar (X11), Sistem Komputer(X12), Simulasi Digital (X13), Perakitan Komputer (X14), Sistem Operasi (X15), Jaringan Dasar (X16) dan Pemograman Web (X17).

Proses clustering menggunakan algoritma K-Means, sebagai berikut :

1. Menentukan jumlah cluster
 Jumlah cluster yang ditentukan untuk mengelompokkan data siswa sebanyak 5cluster.
2. Menentukan centroid atau pusat cluster
 Pusat awal cluster (centroid) ditentukan secara random atau acak, dimana nilai C1 diambil dari baris ke-24, nilai C2 pada baris ke-3, nilai centroid awal pada pengelompokan data nilai rapor siswa dapat dilihat pada tabel 3:

Tabel 3. Centroid Awal

Centroid	Agama	PPKN	B. Indonesia	Matematika	Sejarah	B. Inggris	Seni Budaya	PIUK	PWU	Fikha	Pengalaman Dasar	Sistem Komputer	Simulasi Digital	Peraklan Komputer	Sistem Operasi	Jaringan Dasar	Pemrog. Web
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17
C1	3,34	3,87	3,67	3,00	3,50	3,50	3,34	3,50	3,50	2,84	3,67	3,34	3,33	3,84	3,33	3,50	4,00
C2	2,67	2,67	3,00	2,67	2,67	2,67	2,67	3,00	2,84	2,67	2,67	2,67	3,00	2,67	2,67	3,00	2,67

3. Menghitung jarak dari *centroid*

Menghitung jarak antara titik *centroid* dengan titik tiap objek dengan menggunakan *Euclidian Distance*. Perhitungan *Euclidian Distance* dilakukan dari data 1 sampai data 28.

4. Pengelompokan data

Data dialokasikan ke dalam *centroid* yang paling terdekat. Untuk mengalokasikan kembali data ke dalam masing-masing *cluster* berdasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada, data dialokasikan ke dalam *cluster* yang mempunyai jarak ke *centroid* terdekat dengan data tersebut. Pada tabel 4.berikut ini dapat dilihat perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada.

Tabel 4. Hasil Pengelompokkan Data Iterasi 0

NO	C1	C2	NO	C1	C2
1	2,68	0,83	1	0	1
2	1,74	2,00	2	1	0
3	3,14	0,00	3	0	1
4	2,26	1,42	4	0	1
5	2,05	1,66	5	0	1
6	2,65	1,01	6	0	1
7	1,85	2,01	7	1	0
8	1,47	2,23	8	1	0
9	1,78	1,81	9	1	0
10	2,28	1,33	10	0	1
11	1,92	1,65	11	0	1
12	2,22	1,30	12	0	1
13	2,03	1,63	13	0	1
14	2,30	1,27	14	0	1
15	2,77	0,86	15	0	1
16	2,50	1,33	16	0	1
17	2,60	1,10	17	0	1
18	2,26	1,56	18	0	1
19	2,14	1,55	19	0	1
20	1,90	1,60	20	0	1
21	2,06	1,66	21	0	1
22	1,84	1,67	22	0	1
23	2,23	1,41	23	0	1
24	0,00	3,14	24	1	0
25	2,14	1,31	25	0	1
26	1,29	2,56	26	1	0
27	1,93	1,96	27	1	0
28	2,43	1,12	28	0	1

Dari tabel diatas dapat dilihat kelompok data baru sebagai berikut :

1. Kelompok C1 yaitu data ke : 2, 7, 8, 9, 24, 26 dan 27, ada 7 data pada kelompok C1.
2. Kelompok C2 yaitu data ke : 1, 3, 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 28 ada 21 data pada kelompok C2.

5. Penentuan pusat *cluster* baru

Tentukan posisi centroid baru dengancaramenghitungnilai rata-rata dari data-data yang adapada centroid yang sama. Rajalakshmi [7]

$$\text{Pusat cluster baru} = \frac{\sum_{i=1}^n x_i}{n}; i = 1,2,3,...n \tag{4}$$

Dimana :

x_i : objek ke-i (anggota *cluster*)

n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*.

6. Menghitung jarak dengan pusat *cluster* yang baru.

Iterasi 1 : Sampel data ke 1, data ke 3, data ke 10, data ke 12 dan data ke 24.

Tabel 5. Hasil Pengelompokkan Data Iterasi 1

NO	C1	C2		NO	C1	C2
1	1,57	0,58		1	0	1
2	0,90	1,09		2	1	0
3	2,05	1,13		3	0	1
4	1,23	0,60		4	0	1
5	1,25	1,00		5	0	1
6	1,65	0,64		6	0	1
7	0,96	1,15		7	1	0
8	0,56	1,33		8	1	0
9	0,94	1,23		9	1	0
10	1,23	0,61		10	0	1
11	1,01	0,82		11	0	1
12	1,23	0,74		12	0	1
13	1,14	0,85		13	0	1
14	1,31	0,67		14	0	1
15	1,69	0,62		15	0	1
16	1,47	0,74		16	0	1
17	1,59	0,73		17	0	1
18	1,30	0,72		18	0	1
19	1,24	0,64		19	0	1
20	1,00	0,68		20	0	1
21	1,26	1,05		21	0	1
22	0,97	0,80		22	0	1
23	1,24	0,58		23	0	1
24	1,23	2,20		24	1	0
25	1,28	0,88		25	0	1
26	0,97	1,81		26	1	0
27	1,11	1,26		27	1	0
28	1,36	0,45		28	0	1

Tabel 6. Hasil Pengelompokkan Data Iterasi 0 dan 1.

Iterasi 0			Iterasi 1		
NO	C1	C2	NO	C1	C2
1	0	1	1	0	1
2	1	0	2	1	0
3	0	1	3	0	1
4	0	1	4	0	1
5	0	1	5	0	1
6	0	1	6	0	1
7	1	0	7	1	0
8	1	0	8	1	0
9	1	0	9	1	0
10	0	1	10	0	1
11	0	1	11	0	1
12	0	1	12	0	1
13	0	1	13	0	1
14	0	1	14	0	1
15	0	1	15	0	1
16	0	1	16	0	1
17	0	1	17	0	1
18	0	1	18	0	1
19	0	1	19	0	1
20	0	1	20	0	1
21	0	1	21	0	1
22	0	1	22	0	1
23	0	1	23	0	1
24	1	0	24	1	0
25	0	1	25	0	1
26	1	0	26	1	0
27	1	0	27	1	0
28	0	1	28	0	1

Karena pada iterasi ke-1 posisi cluster sama dengan posisi cluster pada iterasi ke-0 maka proses iterasi dihentikan.

4.1 Hasil dan Analisa :

Dari hasil clustering, maka didapatkan 2 kelompok siswa yaitu siswa:

1. *Cluster 1 (C1)* berjumlah 7 orang siswa, yaitu siswa yang memiliki kompetensi sebagai *artist*.
2. *Cluster 2 (C2)* berjumlah 21 siswa, yaitu siswa yang memiliki kompetensi sebagai *movie makers*.

5. Kesimpulan

Berdasarkan uraian yang sudah dikemukakan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan sebagai berikut :

1. Metode *Clustering* dengan algoritma *K-Means* dapat digunakan untuk mengelompokkan siswa sesuai dengan minatnya pada mata pelajaran produktif dikelas XII berdasarkan nilai rapor saat mereka kelas X.
2. Aplikasi *clustering* yang digunakan yaitu tanagra dapat digunakan untuk membantu sekolah dalam mengelompokkan siswa berdasarkan minatnya yaitu sebagai *artist* atau sebagai *movie makers*.

Daftar Pustaka

- [1] G. Abdillah *et al.*, "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means," vol. 2016, no. Data Mining, pp. 18–19, 2016.
- [2] Afrisawati, "Implementasi Data Mining Pemilihan Pelanggan POTENSIAL

- MENGGUNAKAN ALGORITMA K-MEANS,” *Pelita Inform. Budi Darma*, vol. V, no. 12110955, pp. 157–162, 2013.
- [3] F. E. M. Agustin, A. Fitria, and A. H. S, “(STUDI KASUS : SMP NEGERI 101 JAKARTA) Program Studi Teknik Informatika , Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah,” *Implementasi Algoritm. K-Means Untuk Menentukan Kelompok Pengayaan Mater. Mata Pelajaran Ujian Nas. (Studi Kasus Smp Negeri 101 Jakarta)*, vol. 8, pp. 73–78, 2013.
- [4] E. Muningsih and S. Kiswati, “Penerapan Metode,” *Bianglala Inform.*, vol. 3, no. 1, pp. 229–236, 2015.
- [5] L. R. Angga Ginanjar Maburur, “Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit,” *J. Komput. dan Inform.*, vol. 1, no. 1, pp. 53–57, 2012.
- [6] K. Tampubolon, H. Saragih, B. Reza, K. Epicentrum, A. Asosiasi, and A. Apriori, “Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan,” *Inf. dan Teknol. Ilm.*, pp. 93–106, 2013.
- [7] K. Rajalakshmi, S. S. Dhenakaran, and N. Roobini, “Comparative Analysis of K-Means Algorithm in Disease Prediction,” *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 7, pp. 2697–2699, 2015.

