

APLIKASI PEMETAAN KUALITAS PENDIDIKAN DI INDONESIA MENGUNAKAN METODE K-MEANS

Gibran Satya Nugraha¹, Hairani², Raden Fanny Printi Ardi³

(1) STMIK Bumigora Mataram, (Contact : 081325259291, gibransn@stmikbumigora.ac.id)

(2) STMIK Bumigora Mataram, (Contact : 087839793970, hairani@stmikbumigora.ac.id)

(3) STMIK Bumigora Mataram, (Contact : 087765728722, fanny.printi@stmikbumigora.ac.id)

Abstrak

Aplikasi Pemetaan Kualitas Pendidikan Di Indonesia Menggunakan Metode K-Means adalah sebuah aplikasi yang dirancang untuk memudahkan pemetaan kualitas pendidikan di Indonesia. Aplikasi ini dapat membuat sebuah *cluster* dari kualitas pendidikan di Indonesia berdasarkan sejumlah parameter yaitu Angka Partisipasi Kasar, Angka Partisipasi Murni, Angka Putus Sekolah, Angka Kelulusan, Angka Melanjutkan, Jumlah Sekolah, Rasio Siswa/Sekolah, Rasio Siswa/Kelas, Rasio Kelas/Guru, Rasio Kelas/Sekolah, Rombongan Belajar/Ruang Kelas, dan Jarak Sekolah. Keluaran atau output dari sistem berupa peta yang mengelompokkan daerah-daerah sesuai dengan kualitas pendidikan yang dimilikinya. Analisis perancangan yang digunakan dalam pembuatan sistem ini menggunakan UML (Unified Modeling Language) dimana setiap aktivitas pada sistem akan dikelompokkan secara sendiri-sendiri di dalam sebuah *use case diagram* dan alur dari sistem digambarkan dalam bentuk *flowchart*. Perancangan sistem yang dilakukan antara lain perancangan basis data. Perancangan berdasarkan spesifikasi kebutuhan, dan perancangan antarmuka. Secara umum aplikasi ini dapat menghitung data parameter kualitas pendidikan dengan menggunakan metode K-Means Clustering, dan menampilkan hasilnya dalam bentuk peta, sehingga dinas pendidikan atau lembaga-lembaga yang menangani pendidikan di Indonesia dapat membandingkan kualitas pendidikan setiap provinsi di Indonesia

Key word : Clustering, k-means, pemetaan

I. PENDAHULUAN

Kualitas pendidikan di Indonesia cukup tertinggal dibandingkan dengan negara lain di dunia. Seperti yang dilansir oleh Deutsche Welle, Indonesia berada di peringkat ke lima dari sepuluh negara dalam daftar pemeringkatan kualitas pendidikan di wilayah ASEAN. Kualitas pendidikan di Indonesia tidak lebih baik dari negara Singapura, Brunei Darussalam, Malaysia, dan Thailand. Pada peringkat dunia Indonesia berada di urutan 108, berada di bawah Palestina, Samoa, dan Mongolia [1].

Untuk mengukur kualitas pendidikan di suatu daerah, menurut TNP2K [2] dapat digunakan beberapa parameter atau indikator seperti berikut Penduduk Usia Sekolah, Jumlah Siswa, Angka Partisipasi Kasar, Angka Partisipasi Murni, Angka Putus Sekolah, Angka Kelulusan, Angka Melanjutkan, Jumlah Sekolah, Rasio Siswa/Sekolah, Rasio Siswa/Kelas, Rasio Kelas/Guru, Rasio Kelas/Sekolah, Rombongan

Belajar/Ruang Kelas, dan Jarak Sekolah. Jika nilai-nilai yang diinginkan oleh indikator di atas untuk dapat meningkatkan kualitas pendidikan pada suatu daerah dapat terpenuhi, maka dapat dipastikan daerah tersebut kualitas pendidikannya sudah cukup baik

Peningkatan kualitas pendidikan di suatu negara, tentunya juga harus diawali dengan peningkatan kualitas pendidikan di daerah-daerah dalam negara tersebut, kualitas pendidikan yang baik harus tersebar secara merata di setiap daerah. Kualitas pendidikan pada daerah-daerah Indonesia belum merata dengan maksimal, kurangnya pemerataan pendidikan banyak terjadi di provinsi yang terpencil, hal tersebut banyak disebabkan pembagian subsidi pendidikan yang belum menyeluruh. Untuk menanggulangi hal tersebut dibutuhkan sebuah proses pengelompokan (*clustering*) kualitas pendidikan di tiap provinsi di Indonesia agar dapat diketahui daerah mana saja yang kualitas pendidikannya sudah maju

ataupun masih tertinggal

Untuk membuat proses *clustering* kualitas pendidikan di suatu daerah dapat digunakan sebuah metode yakni K-means *Clustering*. Metode K-Means adalah metode pengelompokan data dengan mengambil parameter sejumlah *k cluster*, dan mempartisi data ke dalam *cluster* tersebut, dengan berpatokan pada kemiripan antar data dalam satu *cluster* dan ketidakmiripan di antara *cluster* yang berbeda, pusat dari *cluster* adalah rata-rata dari nilai anggota *cluster* yang disebut *centroid* atau Center Of Gravity [3].

Berdasarkan latar belakang di atas, maka penulis bermaksud untuk membuat Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means. Proses *clustering* kualitas pendidikan dibuat dengan cara mengolah data parameter atau indikator kualitas pendidikan di Indonesia, dengan menggunakan metode K-means *Clustering*. Kemudian keluaran atau *output* dari hasil proses *clustering* tersebut akan dibuat dalam bentuk peta, sehingga dapat dengan mudah akan dibandingkan kualitas pendidikan antar provinsi di Indonesia, sehingga nantinya pemerataan kualitas pendidikan akan lebih mudah dan cepat.

1.1 Pendidikan

Untuk mengukur kualitas pendidikan di suatu daerah, menurut Chamidi dapat digunakan beberapa indikator [2] sebagai berikut dan untuk contoh data indikator pendidikannya dapat dilihat pada Tabel 1.

1. Angka Partisipasi Kasar (APK), diperoleh dengan membagi jumlah siswa dengan jumlah penduduk menurut kelompok usia sekolah yang sesuai dikalikan 100 persen.
2. Angka Partisipasi Murni, diperoleh dengan membagi jumlah siswa kelompok usia sekolah tertentu dengan jumlah penduduk menurut kelompok usia yang sama dikalikan 100 persen.
3. Angka Putus Sekolah, diperoleh dengan membagi jumlah siswa yang keluar dari sistem pendidikan sebelum lulus selama satu tahun pengajaran tanpa ada surat keterangan pindah dari kepala sekolah dengan jumlah siswa seluruhnya dikalikan 100 persen.
4. Angka Lulusan, diperoleh dengan membagi jumlah siswa yang berhasil menyelesaikan pendidikan untuk suatu jenjang pendidikan

- tertentu dengan jumlah siswa tingkat terakhir pada tahun sebelumnya dikalikan 100 persen.
5. Angka Melanjutkan, diperoleh dengan membagi jumlah siswa baru suatu jenjang pendidikan tertentu dengan jumlah lulusan dari jenjang pendidikan satu tingkat di bawahnya dikalikan 100 persen.
6. Jumlah Sekolah
7. Rasio Siswa/Sekolah, diperoleh dengan membagi jumlah siswa dengan jumlah sekolah pada jenjang pendidikan tertentu.
8. Rasio Siswa/Kelas, diperoleh dengan membagi jumlah siswa dengan jumlah kelas pada jenjang pendidikan tertentu.
9. Rasio Kelas/Guru, diperoleh dengan membagi jumlah kelas dengan jumlah guru pada jenjang pendidikan tertentu .
10. Rasio Kelas/Sekolah, diperoleh dengan membagi jumlah kelas dengan jumlah sekolah pada jenjang pendidikan tertentu.
11. Rombongan Belajar (Rombel)/R. Kelas, diperoleh dengan membagi jumlah rombongan belajar dengan jumlah ruang sekolah pada jenjang pendidikan tertentu.
12. Jarak Sekolah, jarak sekolah disini merupakan rata-rata jarak antar sekolah.

Tabel 1. Indikator Data Pendidikan SD/MI tahun 2010

No	Provinsi	APK (%)	Jarak Sekolah (Km)
1	Nanggroe Aceh Darussalam	109,59	1,63
2	Sumatera Utara	115,53	2,10
3	Sumatera Barat	117,55	0,81
4	Riau	113,99	2,23
5	Jambi	115,15	1,46
6	Sumatera Selatan	113,67	1,88
...
33	Papua	112,66	3,67

1.2 Normalisasi

Menurut Visalakshi et al (2009) normalisasi adalah salah satu proses awal dari sebuah prosedur perhitungan dalam data mining, yang di mana atribut data dibuat dalam bentuk skala

untuk mendapatkan spesifik range yang kecil seperti dari -1.0 sampai 1.0 atau 0.0 sampai 1.0. Normalisasi sebelum proses *clustering* dibutuhkan secara khusus untuk hal-hal yang berkaitan dengan distance metric, seperti Euclidean Distance, yang dimana perbedaannya sangat sensitif di dalam besarnya jarak atau skala dari atribut tersebut [4].

Di dalam aplikasi yang sebenarnya, perbedaan *range* dari sebuah nilai atribut akan menyebabkan satu atribut akan sangat mendominasi atribut lainnya. Normalisasi mencegah dominasi secara sepihak dari sebuah atribut dengan *range* yang besar seperti 'pendapatan' terhadap atribut yang kecil seperti 'umur'. Tujuannya adalah untuk meratakan ukuran atau besarnya jarak dan keragaman dari atribut-atribut lainnya.

Ada banyak metode dalam normalisasi data, seperti Min-Max Normalization, Z-Score Normalization, dan Normalization by Decimal Scaling.

Menurut Subrata et al (2007) Min-Max Normalization adalah nilai dari 'X_{ij}' antar A-v dalam sebuah range (0,1)[5]. Rumus Min-Max Normalization untuk atribut benefit terdapat pada Persamaan (1).

$$r_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad (1)$$

Sedangkan untuk atribut cost menggunakan Persamaan (2).

$$r_{ij} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}} \quad (2)$$

Dengan *r* adalah nilai hasil normalisasi, *x* adalah nilai yang akan dinormalisasika, *i* dan *j* menunjukkan baris dan kolom, serta *min* dan *max* menunjukkan minimum dan maksimum.

Pada Z-Score normalization, nilai dari sebuah atribut A dinormalisasikan berdasarkan rata-rata dan standar deviasi dari A. nilai A yang dilambangkan dengan *v* akan dinormalisasikan terhadap *v'* dengan menggunakan Persamaan (3).

$$v' = \frac{v - \bar{A}}{\sigma_A} \quad (3)$$

Di mana, \bar{A} dan σ_A adalah rata-rata dan standar deviasi secara berurutan dari atribut A. metode normalisasi ini akan sangat berguna ketika nilai minimum dan maksimum secara aktual tidak diketahui.

Normalization by Decimal membuat skala normalisasi dengan cara memindahkan titik desimal dari nilai atribut A. Angka dari titik desimal bergantung dari nilai absolut maksimum dari A. A yang dilambangkan dengan *v* akan dinormalisasikan terhadap *v'* dengan menggunakan Persamaan (4).

$$v' = \frac{v}{10^j} \quad (4)$$

Di mana *j* adalah nilai integer terkecil sehingga $\text{Max}(|v'|) < 1$.

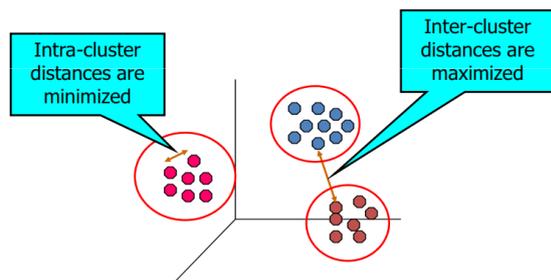
1.3 K-means Clustering

Menurut Efraim et al. (2005) *Clustering* adalah proses pengelompokan objek yang didasarkan pada kesamaan antar objek. Tidak seperti proses klasifikasi yang bersifat *supervised* (terawasi) *learning*, pada *clustering* proses pengelompokan dilakukan atas dasar *unsupervised learning*. Pada proses klasifikasi, akan ditentukan lokasi dari suatu kejadian pada kelas tertentu dari beberapa kelas yang telah teridentifikasi sebelumnya. Sedangkan pada proses *clustering*, proses pengelompokan kejadian dalam kelas akan dilakukan secara alami tanpa mengidentifikasi kelas-kelas sebelumnya [6].

Suatu metode *clustering* dikatakan baik apabila metode tersebut dapat menghasilkan *clusters* dengan kualitas yang sangat baik. Metode tersebut akan menghasilkan *clusters* dengan objek-objek yang memiliki tingkat kesamaan yang cukup tinggi dalam suatu *cluster*, dan memiliki tingkat ketidaksamaan yang cukup tinggi juga apabila objek-objek tersebut terletak pada *cluster* yang berbeda.

Untuk mendapatkan kualitas yang baik, metode *clustering* sangat tergantung pada ukuran kesamaan yang akan digunakan dan kemampuannya untuk menemukan beberapa pola yang tersembunyi. Konsep *clustering*

seperti terlihat pada Gambar 1.



Gambar 1. Konsep clustering

Konsep dasar dari K-Means menurut Efraim *et al.* (2005) adalah pencarian pusat *cluster* secara iteratif. Pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster* [6]. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di-*cluster*, X_{ij} ($i=1, \dots, n$; $j=1, \dots, m$) dengan n adalah jumlah data yang akan di-*cluster* dan m adalah jumlah variabel, serta j merupakan banyaknya parameter yang diuji.

1. Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, K$; $j=1, \dots, m$).
2. Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- i (X_i) pada pusat *cluster* ke- k (C_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean, seperti yang terdapat pada Persamaan (5).

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - C_{kj})^2} \quad (5)$$

3. Suatu data akan menjadi anggota dari *cluster* ke- J apabila jarak data tersebut ke pusat *cluster* ke- J bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya.
4. Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*.
5. Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data yang menjadi anggota pada *cluster* tersebut, dengan menggunakan Persamaan (6)

$$C_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke-}k \quad (6)$$

Ulangi langkah 2-5 hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

Berikut ini adalah dasar algoritma K-Means:

1. Pilih K sebagai centroid awal
2. Ulangi
3. Bentuk K *cluster* dengan menetapkan semua poin ke centroid terdekat.
4. Menghitung centroid setiap *cluster*, sampai centroids tidak berubah

II. METODOLOGI

2.1 Analisis Kebutuhan Sistem

Analisis suatu sistem diperlukan sebagai salah satu proses yang harus dilakukan dalam perancangan dan implementasi suatu sistem. Hal tersebut bertujuan untuk menganalisis, mengevaluasi dan mengetahui kebutuhan suatu perangkat lunak dalam membangun suatu sistem. Aplikasi yang akan dirancang dengan menggunakan Flowchart dan Use Case Diagram

Berdasarkan hasil analisis, maka dapat diketahui kebutuhan yang diperoleh, meliputi kebutuhan masukan sistem, kebutuhan proses, keluaran sistem, kebutuhan perangkat lunak dan kebutuhan perangkat keras.

Untuk jenis *user* yang terlibat dalam sistem ini dibedakan menjadi 4 jenis *user*, yaitu:

1. Admin (*administrator*), adalah *user* dengan hak akses tertinggi pada sistem.
2. Staf (*staff*), adalah *user* yang diberikan (registrasi) hak akses oleh admin untuk dapat menggunakan sistem.
3. Pakar (*expert*), adalah *user* yang diberikan hak (registrasi) oleh admin untuk dapat mengolah data.
4. Pengunjung (*visitor*), adalah *user* yang memiliki hak akses hanya mendapatkan informasi.

Berdasarkan hasil analisis, maka dapat diketahui kebutuhan yang diperoleh, meliputi kebutuhan masukan sistem dan kebutuhan proses.

Untuk jenis *user* yang terlibat dalam sistem ini dibedakan menjadi 4 jenis *user*, yaitu:

1. Admin (*administrator*), adalah *user* dengan hak akses tertinggi pada sistem.
2. Staf (*staff*), adalah *user* yang diberikan (registrasi) hak akses oleh admin untuk dapat menggunakan sistem.
3. Pakar (*expert*), adalah *user* yang diberikan hak (registrasi) oleh admin untuk dapat mengolah data.
4. Pengunjung (*visitor*), adalah *user* yang memiliki hak akses hanya mendapatkan informasi.

Analisis kebutuhan *input* merupakan kebutuhan masukan data yang digunakan untuk aplikasi yang akan dibangun. Masukan bersifat dinamis. Berikut ini adalah rincian masukan sistem (*input*) berdasarkan *user*:

1. *User* admin, data yang dibutuhkan oleh seorang admin antara lain data staf dan pakar
2. *User* Staf, data yang dibutuhkan oleh seorang staf antara lain data provinsi dan kabupaten se-Indonesia
3. *User* Pakar, data yang dibutuhkan oleh seorang pakar antara lain data indikator atau parameter pendidikan tiap provinsi dan kabupaten di Indonesia.

Proses yang akan dibuat adalah sebagai berikut :

1. Admin
 - a. *Login*, proses dimana admin memasukan *username* dan *password* untuk mengakses halaman admin dan mengelola data.
 - b. Manajemen Staf, dimana admin mengelola data staf. Proses manajemen ini dapat memasukan data baru (*input*) dan menghapus data (*delete*).
 - c. Manajemen pakar, dimana admin mengelola data pakar. Proses manajemen ini dapat memasukan data baru (*input*) dan menghapus data (*delete*).
2. Staf
 - a. *Login*, proses dimana staf memasukan *username* dan *password* untuk mengakses halaman staf dan mengelola data. Selain itu staf juga dapat mengubah data diri.
 - b. Manajemen data indikator atau parameter, dimana staf mengelola data indikator-indikator pendidikan di tiap provinsi dan kabupaten. Proses manajemen ini dapat berupa memasukkan data baru (*input*), mengubah data (*edit*) dan menghapus data (*delete*).
 - c. Manajemen data berita, dimana staf mengelola data berita tentang pendidikan. Proses manajemen ini dapat berupa memasukkan data baru (*input*), mengubah data (*edit*) dan menghapus data (*delete*).
 - d. Manajemen data tahun, dimana staf mengelola data tahun. Proses manajemen ini dapat berupa memasukkan data baru (*input*), mengubah data (*edit*) dan menghapus data (*delete*).
 - e. Manajemen data jenjang, dimana staf mengelola data jenjang pendidikan.

Proses manajemen ini dapat berupa memasukkan data baru (*input*), mengubah data (*edit*) dan menghapus data (*delete*).

- f. Manajemen data provinsi dan kabupaten, dimana staf mengelola data provinsi dan kabupaten. Proses manajemen ini dapat berupa memasukkan data baru (*input*), mengubah data (*edit*) dan menghapus data (*delete*).
3. Pakar
 - a. *Login*, proses dimana pakar memasukan *username* dan *password* untuk mengakses halaman pakar dan menganalisis. Selain itu pakar juga dapat mengubah data diri.
 - b. Proses *clustering*, ialah sebuah proses di mana pakar akan melakukan perhitungan K-Means untuk masing-masing data indikator pendidikan yang telah dimasukkan sebelumnya.
 - c. Manajemen analisis, pakar dapat menganalisis data berdasarkan masukan-masukan data dari tiap indikator yang ada. Proses manajemen ini dapat *input* , *edit* , dan hapus hasil analisis.

2.2 Perancangan Sistem

Berdasarkan kebutuhan dapat diketahui apa saja yang menjadi masukan sistem, keluaran sistem, metode yang digunakan sistem, serta antar muka sistem yang akan dibuat, sehingga sistem yang dibuat nantinya sesuai dengan apa yang diharapkan

Menurut Rika et al. (2006) Use Case adalah sekumpulan skenario yang saling terikat untuk mencapai tujuan dari pengguna. Use Case menggambarkan proses sistem dari sudut pandang pengguna [7]. Seorang *actor* adalah sebuah peran yang dijalankan oleh pengguna ketika berinteraksi dengan sistem, seorang *actor* dapat melakukan lebih dari satu Use Case. Gambar 2 merupakan Use Case diagram pada sistem proses *clustering* kualitas pendidikan di Indonesia menggunakan metode K-Means, dengan *actor* yaitu admin, staf, pakar, dan pengunjung. Tugas dari masing-masing *actor* tersebut dapat dilihat pada pembahasan di sub-bab 2.1 Analisis Kebutuhan Sistem



Gambar 2. Use case diagram

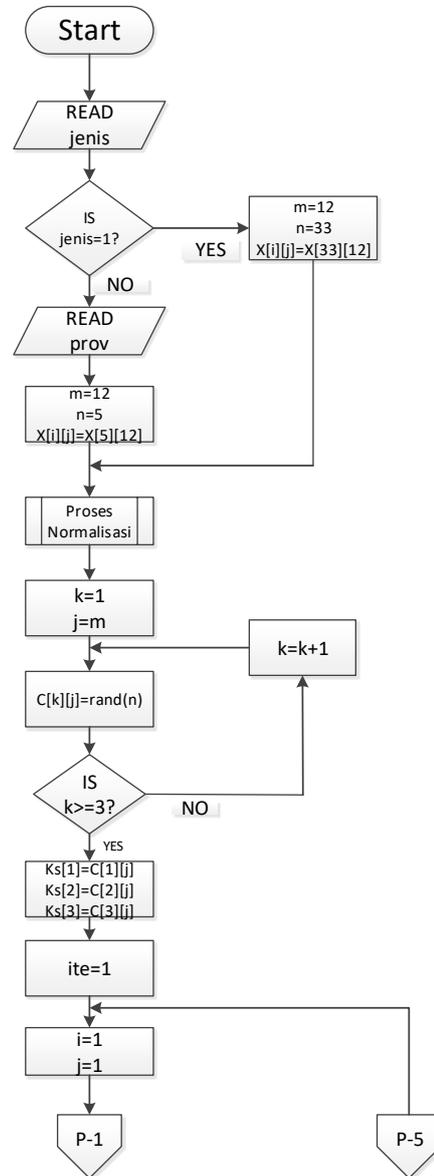
Menurut Nur Afifah (2012) Flowchart adalah gambaran dalam bentuk diagram alir dari algoritma-algoritma dalam program, yang menyatakan arah alur program tersebut. Flowchart ini dibuat pada tahap perancangan program. Fungsinya adalah untuk mengevaluasi jalannya suatu program sebelum direalisasikan dalam bentuk sintaks program sehingga kesalahan alur dalam sintaks program dapat diminimalisir [8].

1. Flowchart proses *clustering* merupakan tahapan-tahapan dalam proses *clustering* kualitas pendidikan dan hasil perhitungan hanya dapat disimpan oleh pakar. Di dalam proses *clustering* ini juga terdapat proses normalisasi. Jumlah kluster yang digunakan ada 3 buah (C1, C2, dan C3), sedangkan untuk jumlah iterasi tidak terbatas. Detail salah satu proses terlihat pada Gambar 3.
2. Flowchart Proses Hitung Pusat Cluster Baru. Proses hitung pusat cluster baru dilakukan jika data yang terdapat pada suatu cluster tidak berpindah lagi ke cluster yang lain pada iterasi setelahnya. Detail proses terlihat pada Gambar 5.
3. Flowchart Proses Normalisasi merupakan tahapan-tahapan dalam proses perhitungan normalisasi nilai indikator proses *clustering* kualitas pendidikan. Detail proses terlihat pada Gambar 6.

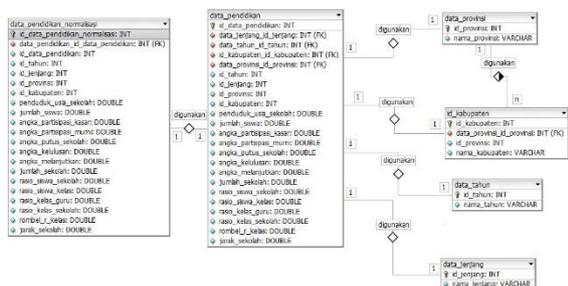
2.3 Perancangan Basis Data

Relasi tabel merupakan hubungan antar tabel pada sistem yang dibangun. Gambar 4 merupakan relasi tabel pada sistem proses

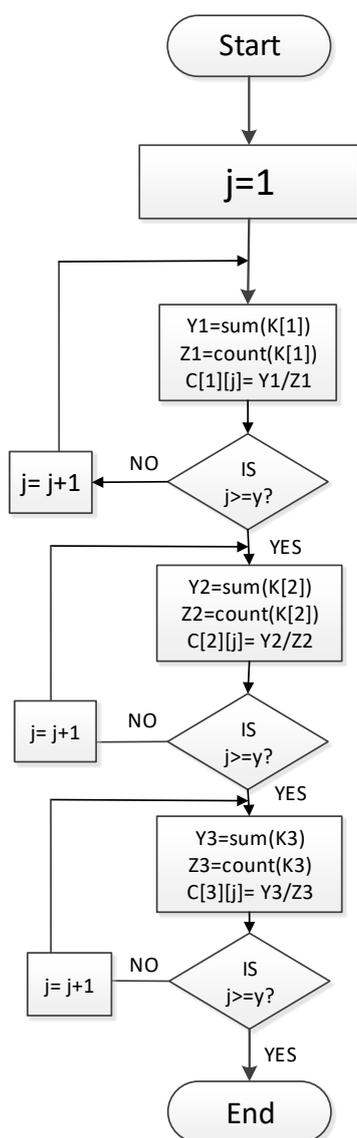
clustering kualitas pendidikan di Indonesia dengan menggunakan metode K-Means lengkap dengan entitas dan kardinalitasnya, yang dimana tabel *data_pendidikan* memiliki relasi dengan tabel *data_tahun*, *data_jenjang*, *data_provinsi*, *data_kabupaten*, *data_pendidikan_normalisasi*.



Gambar 3. Flowchart proses klusterisasi



Gambar 4. Relasi tabel



Gambar 5. Flowchart Proses Hitung Pusat Cluster Baru

III. HASIL DAN PEMBAHASAN

3.1 Implementasi Sistem

Implementasi sistem merupakan tahapan di mana sistem yang telah dirancang sebelumnya dapat berjalan dan dioperasikan. Pada tahapan ini juga berisi tentang penjelasan sistem yang telah dibuat. Berikut ini merupakan penjelasan dari tahapan implementasi sistem.

3.2 Pengujian Sistem

Pengujian sistem merupakan tahapan dimana sistem yang telah dirancang sebelumnya akan diuji pengoperasiannya. Seperti menambah, mengubah, dan menghapus data. Pada tahapan ini juga berisi tentang penjelasan masukan data terhadap sistem yang telah dibuat.

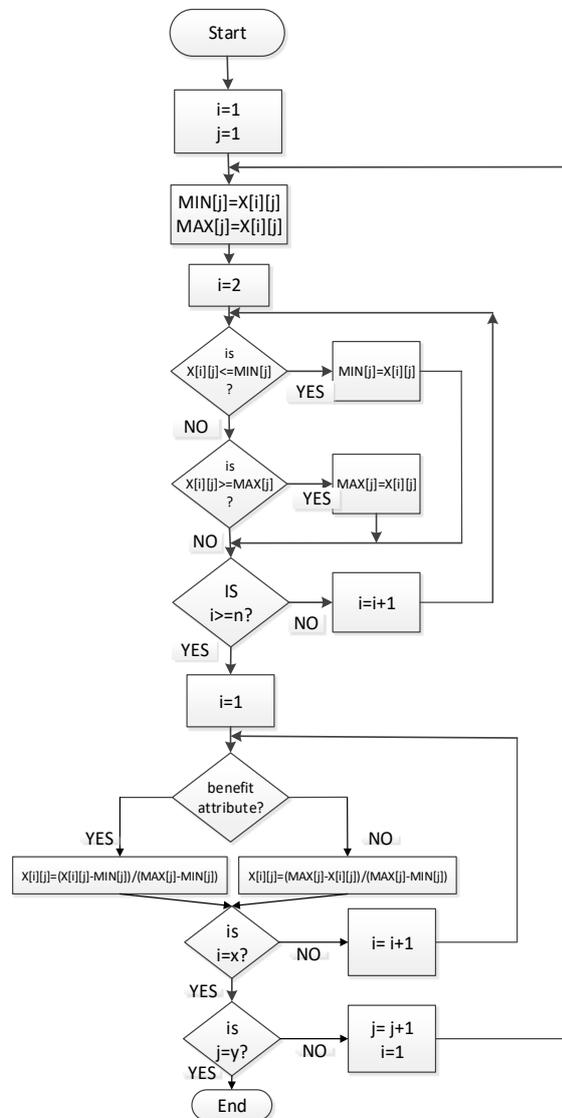
Pengujian sistem proses clustering data pendidikan adalah pengujian yang dilakukan terhadap data yang telah dimasukkan sebelumnya oleh pakar. Data parameter pendidikan akan diolah dalam sebuah perhitungan menggunakan metode K-Means Clustering. Pengujian di sini menggunakan sebuah studi kasus, yaitu menentukan proses clustering kualitas pendidikan di Indonesia pada tahun 2010, dengan jenjang pendidikan SD/MI. Di sini kita akan memilih data pendidikan tahun 2010, dan jenjang pendidikan SD/MI, dapat dilihat pada Tabel 2.

Tabel 2. Data Pendidikan SD/MI Tahun 2010

No	Provinsi	APK (%)	Jarak Sekolah (Km)
1	Nanggroe Aceh Darussalam	109,59	1,63
2	Sumatera Utara	115,53	2,10
3	Sumatera Barat	117,55	0,81
4	Riau	113,99	2,23
5	Jambi	115,15	1,46
6	Sumatera Selatan	113,67	1,88
...
33	Papua	112,66	3,67

Tabel 3 merupakan tabel hasil normalisasi data parameter pendidikan, fungsi dari normalisasi ini ialah untuk menyeragamkan tipe bilangan yang ada, apakah desimal, negatif, atau

sebagainya, dengan melakukan normalisasi maka perhitungan akan lebih akurat. Dalam hal ini penulis menggunakan rumus Min-Max Normalization.



Gambar 6. Flowchart proses normalisasi

Tabel 3. Data Pendidikan Hasil Normalisasi

No	Provinsi	AP K	AP M	...	Jarak Sekolah
1	Nanggroe Aceh Darussalam	0.2	0.2	...	0.61
2	Sumatera Utara	0.67	0.67	...	0.73
3	Sumatera	0.83	0.82	...	0.57

	Barat			...	
4	Riau	0.55	0.55	...	0.76
5	Jambi	0.64	0.64	...	0.6
...
33	Papua	0.43	0.43	...	0.68

Ambil contoh untuk Provinsi Riau dengan parameter pendidikan Angka Partisipasi Kasar (APK) yakni sebesar 113.99 %. Bilangan atau nilai terbesar dan terkecil di kolom data tersebut ialah 119.73 % untuk Provinsi D.K.I Jakarta dan 107.04 % untuk Provinsi Papua Barat secara berurutan. Jika dimasukkan ke dalam rumus, maka akan seperti ini Persamaan (7), Persamaan (7) mengacu pada rumus Min-Max Normalization untuk atribut *benefit* seperti yang terlihat pada Persamaan (1). APK yang diperoleh dengan membagi jumlah siswa dengan jumlah penduduk menurut kelompok usia sekolah yang sesuai kemudian dikalikan 100 persen masuk ke dalam atribut *benefit* karena semakin besar nilai APK di suatu daerah maka semakin berkualitas juga kualitas pendidikan di daerah tersebut

$$r_{ij} = \frac{113.99 - 107.04}{119.73 - 107.04} = 0.55 \quad (7)$$

Hasil perhitungan tersebut ialah 0.55, hasil perhitungan tersebut dapat dilihat di Tabel 3, nilai APK untuk Provinsi Riau.

Untuk melakukan perhitungan proses *clustering*, langkah-langkah yang dilakukan adalah sebagai berikut.

1. Kita harus menentukan titik di setiap *cluster*, menentukan dengan cara sembarang. Adapun titik-titik tersebut ialah:
Cluster 1 = 0.25; 0.40; 0.67; 0.30; 0.35; 0.10; 0.42; 0.14; 0.78; 0.33; 0.44; 0.24
Cluster 2 = 0.48; 0.51; 0.89; 0.62; 0.42; 0.99; 0.88; 0.43; 0.14; 0.84; 0.03; 0.40
Cluster 3 = 0.34; 0.45; 0.55; 0.51; 0.40; 0.54; 0.93; 0.25; 0.29; 0.93; 0.56; 0.57
 Seperti yang kita lihat di atas, untuk masing-masing *cluster* terdapat 12 titik, angka 12 kita dapat dari jumlah parameter pendidikan yang digunakan. Menghitung jarak setiap data terhadap setiap pusat *cluster*, dapat digunakan formula Euclidean, seperti yang

terdapat pada Persamaan (5).

Penjelasan rumus dan algoritma perhitungan sudah dijelaskan di bagian Pendahuluan. Misalnya untuk menghitung jarak data ketiga (Provinsi Sumatera Barat) dengan pusat *cluster* pertama adalah dengan menggunakan perhitungan seperti pada Persamaan (8), parameter yang digunakan masih APK.

$$d_{11} = \sqrt{(0.83-0.25)^2 + (0.82-0.40)^2 + (0.59-0.67)^2 + (0.44-0.30)^2 + \sqrt{(0.44-0.35)^2 + (0.18-0.10)^2 + (0.57-0.42)^2 + (0.42-0.14)^2 + \sqrt{(0.14-0.78)^2 + (0.41-0.33)^2 + (0.29-0.44)^2 + (0.02-0.24)^2} = 1.0689 \quad (8)$$

Kemudian menghitung jarak data data ketiga (Provinsi Sumatera Barat) dengan pusat *cluster* kedua menggunakan perhitungan seperti pada Persamaan (9)

$$d_{12} = \sqrt{(0.83-0.48)^2 + (0.82-0.51)^2 + (0.59-0.89)^2 + (0.44-0.62)^2 + \sqrt{(0.44-0.42)^2 + (0.18-0.99)^2 + (0.57-0.88)^2 + (0.42-0.43)^2 + \sqrt{(0.14-0.14)^2 + (0.41-0.84)^2 + (0.29-0.03)^2 + (0.02-0.40)^2} = 1.182 \quad (9)$$

Terakhir adalah menghitung jarak data ketiga (Provinsi Sumatera Barat) dengan pusat *cluster* ketiga menggunakan perhitungan seperti pada Persamaan (10)

$$d_{13} = \sqrt{(0.83-0.34)^2 + (0.82-0.45)^2 + (0.59-0.55)^2 + (0.44-0.51)^2 + \sqrt{(0.44-0.40)^2 + (0.18-0.54)^2 + (0.57-0.93)^2 + (0.42-0.25)^2 + \sqrt{(0.14-0.29)^2 + (0.41-0.93)^2 + (0.29-0.56)^2 + (0.02-0.57)^2} = 1.036 \quad (10)$$

- Untuk data provinsi yang lain menggunakan perhitungan yang sama dengan menggunakan Persamaan (8), (9), dan (10). Selanjutnya akan dibandingkan nilai dari tiap *cluster* pada masing-masing provinsi, jika suatu provinsi pada *cluster* 1 memiliki jarak yang terkecil dengan pusat *cluster*, dibandingkan dengan nilai yang ada pada *cluster* 1 dan 2 maka otomatis provinsi tersebut akan berada di *cluster* 1, apabila jarak terkecil berada di *cluster* 2 maka provinsi tersebut ada di *cluster* 2, begitu pula dengan di *cluster* 3.
- Hasil perhitungan akhir dari iterasi dapat dilihat di Tabel 4 baris pertama, di sana

terlihat Provinsi Nanggroe Aceh Darussalam memiliki jarak yang terkecil dengan pusat *cluster* yakni pada *cluster* pertama (0.71), Provinsi Sumatera Utara pada *cluster* ketiga (0.51), untuk Provinsi Sumatera Barat terletak pada *cluster* pertama (0.46), begitu seterusnya untuk provinsi lainnya. Sekarang kita lihat 3 kolom terakhir (C1, C2, C3), fungsi ketiga kolom tersebut untuk membantu dalam proses menentukan suatu provinsi berada di *cluster* mana, ketiga nilai pada kolom Cluster 1, 2, dan 3 akan dibandingkan. Nilai terkecil akan diberikan simbol angka 1, dan yang lainnya mendapatkan simbol angka 0.

- Tabel pusat *cluster* masing-masing parameter kualitas pendidikan terhadap *cluster* 1, 2, dan 3 dapat dilihat pada Tabel 5.
- Setelah mengetahui hasil *clustering* masing-masing provinsi, maka pakar akan menentukan *cluster* yang tepat sesuai dengan analisis pakar itu sendiri (lihat Gambar 7). Pada Gambar 7 tersebut terdapat daftar kelompok provinsi, jumlah provinsi yang berada pada suatu *cluster*, jumlah rata-rata nilai pada masing-masing *cluster*, serta *cluster* yang akan dipilih oleh pakar sesuai dengan analisa pakar tersebut.

Tabel 4. Data pendidikan hasil perhitungan K-means.

Provinsi	Cluster (nilai)			C1	C2	C3
	1	2	3			
NAD	0.71	1.56	1.19	1	0	0
Sumatera Utara	0.77	0.8	0.51	0	0	1
Sumatera Barat	0.46	1.14	0.84	1	0	0
Riau	0.44	1.31	0.75	1	0	0
Jambi	0.63	1.36	0.96	1	0	0
...
Papua	0.82	1.27	0.75	0	0	1

Tabel 5. Data pusat *cluster* tiap parameter

Parameter	Cluster		
	1	2	3
Angka Partisipasi Kasar	0.721	0.385	0.885
Angka Partisipasi Murni	0.71	0.383	0.875
Putus Sekolah	0.656	0.365	0.595

Angka Kelulusan	0.551	0.608	0.61
Angka Melanjutkan	0.512	0.83	0.19
Jumlah Sekolah	0.224	0.073	0.09
Rasio Siswa/Sekolah	0.625	0.565	0.41
Rasio Siswa/Kelas	0.487	0.473	0.025
Rasio Siswa/Guru	0.232	0.303	0.87
Rasio Kelas/Sekolah	0.491	0.345	0.495
Rombel/R. Kelas	0.406	0.475	0.755
Jarak Sekolah	0.923	0.08	0.75

menganalisis penggolongan *cluster* yang tepat berdasarkan data masukan parameter pendidikan, dapat dilihat pada Gambar 9, sedangkan untuk hasil analisisnya dapat dilihat pada Gambar 10.

Nama Provinsi	Jumlah	Rata-rata Nilai Pusat Cluster	Cluster
Bali, Banten, DI Yogyakarta, DKI Jakarta, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kep. Bangka Belitung, Aceh, Nusa Tenggara Barat, Sulawesi Barat, Sulawesi Selatan, Sulawesi Utara, Riau, Sumatera Barat, Kep. Riau, Lampung, Nusa Tenggara Timur, Sulawesi Tenggara, Kalimantan Selatan, Bengkulu, Sumatera Selatan, Gorontalo, Sulawesi Tengah, Sumatera Utara	27	0.545	Cluster 2
Papua Barat, Maluku, Kalimantan Timur, Papua	4	0.546	Cluster 3
Maluku Utara, Kalimantan Tengah	2	0.407	Cluster 1

Gambar 7. Data kualitas pendidikan masing-masing cluster

6. Setelah pakar memilih *cluster* yang tepat untuk masing-masing kelompok provinsi maka akan ditampilkan peta hasil *clustering* kualitas pendidikan. Hasil *clustering* dalam bentuk peta tersebut dapat dilihat pada Gambar 8, di sana terlihat provinsi yang berada di *cluster* satu (Maluku Utara dan Kalimantan Tengah) berwarna biru, untuk provinsi yang berada di *cluster* dua (Bali, Banten, DI. Yogyakarta, DKI Jakarta, dll) akan berwarna coklat, dan untuk provinsi yang berada di *cluster* tiga (Papua Barat, Maluku, Kalimantan Timur, dan Papua) akan berwarna hijau, pembagian warna tersebut dapat dilihat pada legenda



Gambar 8. Hasil proses clustering kualitas pendidikan dalam bentuk peta

7. Setelah peta kualitas pendidikan telah didapatkan, selanjutnya adalah pakar akan



Gambar 9. Proses analisis data pendidikan oleh pakar

Data Kualitas Pendidikan SD/MI di Indonesia, Tahun 2010



Gambar 10. Hasil analisis data kualitas pendidikan

Perhitungan untuk data kabupaten juga bisa dilakukan, dengan menggunakan langkah yang sama pada penjelasan sebelumnya mengenai pemetaan kualitas pendidikan untuk provinsi di seluruh Indonesia. Sebagai contoh dalam perhitungan ini penulis menggunakan data dari Provinsi Yogyakarta, data indikator pendidikannya dapat dilihat pada Tabel 6. Untuk hasil pemetaan kualitas pendidikannya dapat dilihat pada Gambar 11. Dari hasil pemetaan terlihat bahwa untuk *cluster* 1 terdiri dari Kota Yogyakarta dan Kabupaten Bantul, sedangkan untuk *cluster* 2 terdiri dari Kabupaten Sleman saja, dan yang terakhir *cluster* 3 terdiri dari Kabupaten Kulon Progo, dan Gunung Kidul.

Tabel 6. Data indikator pendidikan SMA Provinsi Yogyakarta tahun 2010

No	Kabupaten	APK (%)	Jarak Sekolah (Km)
1	Kulon Progo	109,59	1,63
2	Bantul	115,53	2,10

3	Gunung Kidul	117,55	0,81
4	Sleman	113,99	2,23
5	Yogyakarta	115,15	1,46



Gambar 11. Pemetaan kualitas pendidikan Provinsi Yogyakarta

IV. SIMPULAN DAN SARAN

4.1 Simpulan

Berdasarkan analisis terhadap Aplikasi Pemetaan Kualitas Pendidikan di Indonesia, maka didapatkan beberapa kesimpulan sebagai berikut:

1. Metode K-Means Clustering dapat digunakan dalam membuat pemetaan kualitas pendidikan di Indonesia
2. Hasil proses *clustering* dibuat dalam bentuk peta sehingga dinas pendidikan atau lembaga-lembaga yang menangani pendidikan di Indonesia dapat membandingkan kualitas pendidikan setiap provinsi atau kabupaten di Indonesia

4.2 Saran

Beberapa saran untuk pengembangan Aplikasi Pemetaan Kualitas Pendidikan di Indonesia ini antara lain sebagai berikut

1. Kriteria-kriteria yang digunakan sebagai parameter pendidikan dapat ditambah secara dinamis
2. Perlu adanya analisis jumlah cluster yang optimal sebelum dilakukan proses *clustering*

REFERENSI

- [1] A. Sahroji, "Daftar Negara ASEAN dengan Peringkat Pendidikan Tertinggi," 2017. [Online]. Available: <https://news.okezone.com/read/2017/11/24/18/1820178/daftar-negara-asean-dengan->

- peringkat-pendidikan-tertinggi. [Accessed: 12-Apr-2018].
- [2] C. Shafrudin, Makna dan Aplikasi Sederhana Indikator Pendidikan. 2005. Departemen Pendidikan Nasional, Badan Pendidikan dan Pengembangan, Pusat Data dan Informasi Pendidikan, Bidang Pendayagunaan Data dan Informasi, Jakarta.
- [3] H. Kamber. 2007. Data Mining Concepts and Techniques Second Edition. Elsevier Inc.
- [4] N. K. Visalakshi and K. Thangavel, "Impact of normalization in distributed K-means clustering," International Journal of Soft Computing, vol. 4, no. 4, 2009, pp. 168–172.
- [5] S. Chakraborty, "A Simulation Based Comparative Study of Normalization Procedures in Multiattribute Decision Making," 6th WSEAS Int. Conf. Artif. Intell. Knowl. Eng. Data Bases, vol. Proceeding, 2007, pp. 102–109.
- [6] E. Turban, J. E. Aronson, and T.-P. Liang. 2005. Decision Support Systems and Intelligent Systems. International Edition, 7th Edition, Pearson Prentice-Hall Education International, New Jersey.
- [7] R. Michael Yoseph and R. Michael Yoseph. 2006. "Analisis dan Perancangan Sistem Informasi Laboratorium Rumah Sakit Kanker Dharmais dengan menggunakan Total Architecture Syntesis," Binus University Jakarta.
- [8] N. Afifah. 2012. "Sistem Pendukung Keputusan Penerimaan Beasiswa Magang Menggunakan Metode SAW (Simple Additive Weighting)," Universitas Trunojoyo Madura.