



# Implementasi Metode *K-Means* berbasis *Chi-Square* pada Sistem Pendukung Keputusan untuk Identifikasi Disparitas Kebutuhan Guru

M. Nishom\*, Dega Suroño Wibowo

<sup>a,b</sup> Politeknik Harapan Bersama Tegal

Naskah Diterima : 20 September 2018; Diterima Publikasi : 1 November 2018

DOI : 10.21456/vol8iss2pp187-194

## Abstract

In this research, the Chi-Square-based K-Means method was implemented in the Decision Support System (DSS) to identify the disparity in Teacher's needs compared to the real conditions of Teacher's availability in the education unit (school). This is very important, because based on data from the UNESCO Institute for Statistics shows that the ratio between teachers and students in Indonesia is the lowest in the world. This is influenced by the distribution of Teachers who do not meet the needs and exceed the number of student enrollments at all levels of education, resulting in less optimal quality of education produced in various regions in Indonesia. Thus, it is necessary to group data and label the disparity of Teacher's needs in educational units in various regions in Indonesia, especially in the Tegal City. In this case, the K-Means Clustering method was used to group data based on Teacher's availability data, and Chi-Square analysis was used to determine the disparity in Teacher's needs with the condition of Teacher's availability. Data collection methods used in research are observation methods. The results showed that the DSS application that had been produced could dynamically determine the education unit cluster based on the teacher availability disparity category in the Tegal City. In addition, labeling of the K-Means cluster based on the Chi-square test has a high degree of accuracy, which is 84.47%.

**Keywords** : Clustering; K-Means; Chi-Square; Teacher Disparity; Student-Teacher Ratio.

## Abstrak

Pada penelitian ini, metode *K-Means* berbasis *Chi-Square* diimplementasikan pada Sistem Pendukung Keputusan (SPK) untuk mengidentifikasi disparitas kebutuhan Guru dibandingkan dengan kondisi riil ketersediaan Guru di satuan pendidikan (sekolah). Hal ini sangat penting, karena berdasarkan data dari *UNESCO Institute for Statistics* menunjukkan bahwa rasio perbandingan antara guru dan murid di Indonesia merupakan yang terendah di dunia. Hal itu dipengaruhi oleh persebaran Guru yang tidak sesuai dengan kebutuhan dan melampaui jumlah pendaftaran murid di segala tingkat pendidikan, sehingga mengakibatkan kurang optimalnya kualitas pendidikan yang dihasilkan di berbagai wilayah di Indonesia. Dengan demikian, perlu dilakukan pengelompokan data dan pelabelan pada disparitas kebutuhan Guru pada satuan pendidikan di berbagai wilayah di Indonesia, khususnya di wilayah Kota Tegal. Pada kasus ini, metode *K-Means Clustering* digunakan untuk mengelompokkan data berdasarkan data ketersediaan Guru, dan analisa *Chi-Square* digunakan untuk mengetahui disparitas kebutuhan Guru dengan kondisi ketersediaan Guru. Metode pengumpulan data yang digunakan dalam penelitian adalah metode observasi. Hasil penelitian menunjukkan, bahwa aplikasi SPK yang telah dihasilkan dapat secara dinamis menentukan klaster satuan pendidikan berdasarkan kategori disparitas ketersediaan Guru di wilayah Kota Tegal. Selain itu, pelabelan dari klaster *K-Means* berdasarkan uji *Chi-square* memiliki tingkat akurasi yang tinggi, yaitu 84.47%.

**Kata kunci**: Klasterisasi; K-Means; Chi-Square; Disparitas Guru; Rasio Guru-Murid.

\*) Penulis korespondensi: nishom@poltektegal.ac.id

## 1. Pendahuluan

Salah satu faktor yang sangat penting dalam upaya pembangunan pendidikan di Indonesia adalah ketersediaan guru yang memadai, baik secara kualitas maupun kuantitas. Ketersediaan guru yang memadai dihadapkan pada dua masalah pokok, yaitu pemenuhan kebutuhan tenaga guru yang belum sesuai dengan kebutuhan daerah dan peningkatan kualitas profesional yang belum memenuhi standar minimal. Kedua permasalahan inilah yang pada akhirnya menimbulkan terjadinya disparitas yang tinggi terhadap angka kebutuhan guru di berbagai wilayah di Indonesia (Sunuyeko *et al.*, 2017).

Persebaran Sumber Daya Manusia (SDM) pendidik yang tidak merata di Indonesia saat ini mengakibatkan dampak negatif bagi kualitas pendidikan yang dihasilkan di berbagai wilayah di Indonesia. Hal tersebut dapat dilihat dari rasio antara guru dan murid yang sangat rendah seperti disampaikan oleh institut statistik UNESCO, yaitu mencapai 17:1. Penyebab dari permasalahan tersebut adalah terpusatnya tenaga pendidik (guru) di kota-kota besar dan tidak terpenuhinya kebutuhan tenaga pendidik di daerah perbatasan, daerah tertinggal, daerah kepulauan, dan daerah-daerah terpencil. Sedangkan peraturan Pemerintah Nomor 74 Tahun 2008 tentang Guru menyebutkan bahwa guru memiliki beban kerja paling sedikit 24 (dua puluh empat) jam tatap muka dan sebanyak-banyaknya 40 (empat puluh) jam tatap muka per minggu. Untuk itu, analisa kebutuhan tenaga pendidik sangat diperlukan di Indonesia.

Terdapat beberapa metode penentuan kebutuhan tenaga pendidik menurut Pusat Data dan Statistik Pendidikan dan Kebudayaan (PDSPK), yaitu studi dokumentasi statistik berdasarkan persekolahan dan data pokok pendidikan yang diolah kembali berdasarkan item data kepala sekolah dan guru dengan melihat perbandingan rasio siswa dan guru. Hasil analisis tersebut menunjukkan bahwa di Indonesia masih diperlukan tambahan tenaga pendidik sebesar 146.957 tenaga pendidik (Kemendikbud, 2016). Namun, angka tersebut tidak dapat menunjukkan kelompok atau *cluster* kebutuhan SDM pendidik pada masing-masing wilayah di Indonesia, sehingga proyeksi kebutuhan normatif SDM pendidik untuk masing-masing wilayah tidak dapat diperoleh. Selain itu, evaluasi terhadap metode yang digunakan untuk menentukan kebutuhan SDM pendidik juga seharusnya diperhatikan, terutama terkait tingkat akurasi.

Salah satu algoritma data mining yang banyak digunakan dalam penelitian adalah *K-Means* (Wu *et al.*, 2014), karena algoritma ini masuk pada 10 besar dalam metode data mining yang paling sering digunakan (Wu *et al.*, 2008).

*K-Means* merupakan salah satu jenis algoritma *clustering* non-hirarki yang dapat digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan antar datanya. Salah satu hal yang membuat algoritma ini menjadi populer adalah kemudahan untuk mengimplementasikan dalam berbagai jenis kasus dan atribut data. Penelitian terkait penggunaan metode *K-Means* telah dilakukan untuk klasterisasi angkatan kerja (Kuswantoro *et al.*, 2015). Penerapan metode *K-Means clustering* ini dapat digunakan untuk mengelompokkan data ketersediaan Sumber Daya Manusia pendidik (guru) pada suatu wilayah ke dalam sejumlah *cluster* sesuai dengan tingkat kemiripan dari data-datanya.

Berdasarkan kondisi dan permasalahan yang telah disebutkan, penelitian ini bertujuan membangun aplikasi dengan mengimplementasikan metode *K-Means* berbasis hasil uji *Chi-Square* pada Sistem Pendukung Keputusan untuk mengidentifikasi profil disparitas kebutuhan Guru dibandingkan dengan kondisi nyata (riil) ketersediaan Guru di satuan pendidikan.

## 2. Kerangka Teori

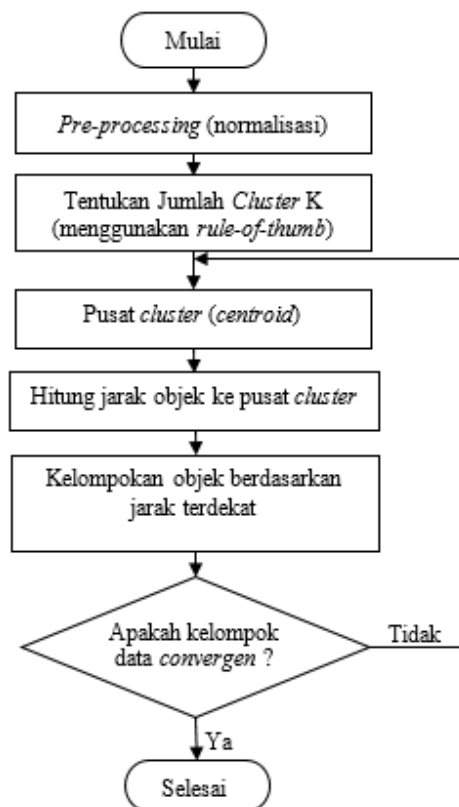
### 2.1. Clustering

Pengelompokan data atau biasa disebut *clustering* merupakan suatu cara atau metode yang biasa digunakan untuk mengklaster atau mengelompokkan kumpulan suatu data ke dalam sebuah himpunan atau klaster, mencari dan mengelompokkan data yang memiliki kesamaan atau kemiripan dalam sifat atau ciri-ciri (*similarity*) antara data satu dengan data yang lainnya dalam sebuah dataset. Sifat dari metode ini adalah *unsupervised* atau tanpa arahan (artinya metode ini diimplementasikan tanpa adanya *training* atau latihan dan tanpa ada *teacher*) serta tidak memerlukan target atau sasaran luaran. Tujuan algoritma klaster adalah menciptakan klaster yang koheren secara internal, tetapi jelas berbeda satu sama lain. Dengan kata lain, data dalam sebuah klaster harus semirip mungkin dan data dalam satu klaster harus sebeda mungkin dari data dalam klaster lainnya (Manning, 2008).

### 2.2. K-Means Clustering

Algoritma *K-means* merupakan suatu algoritma yang biasa digunakan untuk menemukan kelompok dari dokumen atau objek yang *non-overlapping*. Algoritma klaster *K-Means* juga disebut sebagai suatu algoritma yang sangat efektif untuk mengelompokkan kumpulan data (Larose, 2014). Algoritma *K-Means clustering* juga sangat sederhana untuk diimplementasikan dan dijalankan, proses *clustering* relatif cepat, mudah beradaptasi serta umum penggunaannya dalam praktek (Prasetyo, 2014). Proses algoritma ini ide atau alurnya cukup sederhana. Di tahap awal, terlebih dahulu ditentukan

jumlah kelompok atau cluster yang akan dipakai. Kemudian dilanjutkan dengan memilih dokumen pertama atau elemen pertama dalam sebuah *cluster* untuk digunakan sebagai *cluster centroid point* (titik tengah klaster). Selanjutnya, dilakukan iterasi atau pengulangan langkah-langkah dalam menentukan jarak dokumen atau objek dengan centroid, sampai terjadi kestabilan dan semua kelompok objek telah konvergen (Oyelade *et al.*, 2010). Adapun tahapan dalam algoritma *K-Means* dengan implementasi pendekatan *rule-of-thumb* ditunjukkan pada Gambar 1 di bawah ini.



Gambar 1. Flowchart Algoritma *K-Means*

Pertama, sebelum masuk ke proses *clustering*, diperlukan analisis terhadap tipe data yang ada, apakah data tersebut perlu untuk dinormalisasi atau tidak. Misalnya, terkait pencatatan tingkat atau jumlah kematian dari data penduduk di negara Indonesia pada setiap bulan yang didasarkan pada jenis usia atau umur. Biasanya, akan terdapat tiga dimensi atau kelompok data, yaitu jumlah kematian 0-jutaan, penduduk usia 1 sampai 12 bulan, dan umur 0 sampai 100 tahun. Jika jarak dari masing-masing dimensi dibentangkan, maka akan terjadi ketidakseimbangan pada dimensi ketiga (jumlah kematian). Dalam kasus ini, maka diperlukan normalisasi data sebelum proses klastering. Normalisasi yang dapat digunakan adalah normalisasi *Min-Max*. Cara ini diimplementasikan dengan merubah nilai atau data asli ke bentuk linier dilakukan dengan menggunakan persamaan berikut:

$$x' = \frac{x - nilai_{min}}{nilai_{max} - nilai_{min}} \quad (1)$$

dimana,  $x$  = data per kolom,  $nilai_{min}$  = nilai terkecil dari data per kolom,  $nilai_{max}$  = nilai terbesar dari data perkolom. Kedua, menentukan jumlah *cluster* ( $K$ ). Proses ini menginisialisasi nilai awal  $K$  sebagai jumlah cluster yang akan dipartisi secara dinamis. Dalam menentukan jumlah  $K$  digunakan pendekatan *rule-of-thumb* dengan menggunakan persamaan sebagai berikut:

$$k = \sqrt{\frac{n}{2}} \quad (2)$$

dimana  $n$  = jumlah objek yang akan dikelompokkan, dan  $k$  = jumlah cluster

Ketiga, menentukan *centroid* awal (initial centroid). Banyak metode yang dapat digunakan, seperti dengan metode *random* (mengambil secara acak). Sedangkan untuk menentukan *centroid* baru, dapat dilakukan dengan menghitung nilai rata-rata dari total nilai objek dalam *cluster* baru. Rumus yang digunakan adalah sebagai berikut:

$$C_i = \frac{\sum_{i=1}^n x_i \in s_i}{n} \quad (3)$$

dimana  $C_i$  = *centroid* baru ke  $i$ ,  $s_i$  = objek ke  $i$ ,  $x_i$  = nilai pada objek ke  $i$ ,  $n$  = jumlah data pada tiap kelompok.

Keempat, menghitung jarak objek dengan *centroid*. Untuk menghitung jarak antara objek dengan pusat *cluster* (*centroid*) dapat dilakukan dengan menggunakan beberapa pendekatan. Pada penelitian ini digunakan rumus *Euclidean Distance*. Perhitungan ini menghitung nilai kuantitatif dari tingkat kemiripan atau ketidakmiripan data (*proximity measure*) yang dapat menghasilkan jarak dari objek dengan pusat *cluster* (*centroid*). Berikut merupakan rumus *Euclidean Distance* yang digunakan untuk menghitung jarak objek dengan pusat *cluster*:

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

dimana,  $d$  = jarak antara  $x$  dan  $y$ ,  $x$  = data pusat klaster,  $y$  = data pada atribut,  $i$  = setiap data,  $n$  = jumlah data,  $x_i$  = data pada pusat klaster ke  $i$ ,  $y_i$  = data pada setiap data ke  $i$ .

Kelima, pengelompokan objek berdasarkan jarak terdekat. Sebelum pengelompokan objek dilakukan, pertama harus dilakukan perhitungan untuk menentukan jarak yang bernilai minimum.

Setelah didapatkan nilai minimum, dilakukan pengelompokan objek. Tahapan akhir adalah melakukan uji konvergensi antara kelompok data baru dan kelompok data pada proses sebelumnya, jika kelompok data yang baru konvergen dengan kelompok data sebelumnya, maka proses clustering selesai. Jika tidak, maka lakukan iterasi dimulai dari penentuan pusat *cluster* baru.

### 2.3. Uji Homogenitas Klaster

Pengujian homogenitas klaster dapat ditentukan berdasarkan nilai koefisien *silhouette* yang dapat diperoleh melalui beberapa tahapan. Pertama, Hitung rata-rata jarak dari suatu objek, misalkan  $i$  dengan semua objek lain yang berada dalam satu klaster dengan menggunakan persamaan di bawah ini.

$$a_i = \frac{1}{|A| - 1} \sum_{j \in A, i \neq j} d(i, j) \quad (5)$$

dimana,

$|A|$  = banyaknya data dalam klaster  $A$ , dan  $i, j$  = indeks dari dokumen, sedangkan  $d(i, j)$  = jarak antara dokumen ke  $i$  dengan dokumen ke  $j$ .

Kedua, hitung rata-rata jarak dari dokumen  $i$  tersebut dengan semua dokumen dalam klaster lain, dan diambil nilai terkecilnya.

$$d(i, c) = \frac{1}{|A|} \sum_{j \in c} d(i, j) \quad (6)$$

dimana,  $d(i, c)$  adalah jarak rata-rata objek  $i$  dengan semua objek pada klaster lain  $C$  dimana  $A \neq C$ .

$$b(i) = \min_{C \neq A} d(i, C) \quad (7)$$

Ketiga, hitung nilai *silhouette coefficient*-Nya dengan persamaan berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

### 2.3. Chi-Square

Kai kuadrat atau sering disebut dengan *Chi-square* merupakan sebuah uji hipotesis tentang perbandingan antara frekuensi observasi dengan frekuensi harapan yang didasarkan oleh hipotesis tertentu pada setiap kasus atau data yang diambil untuk diamati. Pengujian ini sangat bermanfaat dalam melakukan analisis statistik jika kita tidak memiliki informasi tentang populasi atau jika asumsi-asumsi yang dipersyaratkan untuk penggunaan statistik parametrik tidak terpenuhi. Pada penelitian ini, nilai interpretasi dari *Chi-Square* akan didasarkan pada

perbandingan antara nilai-nilai aritmatika *Chi-Square* dengan nilai tabel *Chi-Square*. Nilai yang lebih besar dari aritmatika *Chi-Square* berarti ada perbedaan yang signifikan antara kondisi ketersediaan tenaga pengajar dan kebutuhan normatif tenaga pengajar, begitu juga sebaliknya (Stine, 2011). Persamaan yang dapat digunakan adalah seperti berikut:

$$X^2 = \sum \frac{(F_o - F_h)^2}{F_h} \quad (9)$$

dimana,  $X^2$  = Nilai *Chi-square*,  $F_o$  = Nilai yang diobservasi, dan  $F_h$  = Nilai yang diharapkan.

### 2.5. Analisa Sumber Daya Manusia Pendidik (Guru)

Menurut Undang-undang RI Nomor 20 Tahun 2003, Tenaga Pendidik merupakan tenaga kependidikan yang berkualifikasi sebagai guru, dosen, konselor, pamong belajar, widyaiswara, tutor, instruktur, fasilitator dan sebutan lain yang sesuai dengan kekhususannya, serta berpartisipasi dalam penyelenggaraan pendidikan. Pendidik mempunyai dua arti, yaitu arti yang luas dan arti yang sempit. Dalam arti luas, seorang pendidik adalah semua orang yang berkewajiban membina peserta didik. Dalam arti sempit, pendidik merupakan orang yang dengan sengaja dipersiapkan menjadi guru atau dosen. Guru dan dosen adalah jabatan profesional, sebab mereka mendapatkan tujangan profesional.

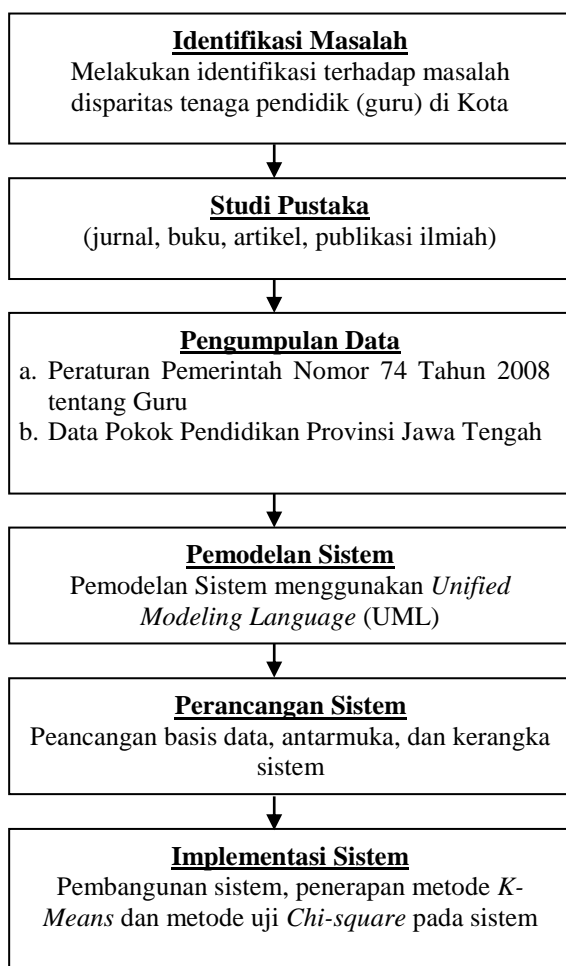
Metode yang diterapkan untuk merencanakan kebutuhan tenaga pendidik (guru) di sekolah tingkat dasar dan menengah adalah "rasio guru-siswa", yaitu rasio guru terhadap jumlah siswa untuk suatu satuan pendidikan berdasarkan tingkat satuan pendidikan. Standar rasio guru-siswa ini ditentukan berdasarkan Peraturan Pemerintah Nomor 74 Tahun 2008 tentang guru pasal 17, bahwa rasio guru-murid yang ideal adalah sebagai berikut: untuk satuan pendidikan TK dan RA, atau yang sederajat adalah 15:1, SD atau yang sederajat adalah 20:1, MI atau yang sederajat adalah 15:1, SMP atau yang sederajat adalah 20:1, MTs atau yang sederajat adalah 15:1, SMA atau yang sederajat adalah 20:1, MA atau yang sederajat adalah 15:1, SMK atau yang sederajat adalah 15:1, dan MAK atau yang sederajat adalah 12:1. *Chi-square* merupakan salah satu analisa disparitas yang dapat digunakan untuk menentukan kesesuaian antara ketersediaan SDM pendidik dengan standar "rasio siswa-guru". Interpretasi terhadap nilai *Chi-square* adalah berdasarkan perbandingan antara nilai *Chi-square* hitung dengan nilai *Chi-square* tabel, yakni apabila nilai *Chi-square* hitung lebih besar maka terdapat perbedaan signifikan antara eksistensi tenaga pendidik (guru) dengan kebutuhan normatif tenaga pendidik dan berlaku sebaliknya.

### 3. Metode

Penelitian ini menggunakan metode *K-Means* untuk melakukan klusterisasi satuan pendidikan di wilayah Kota Tegal. Sedangkan untuk menguji tingkat akurasi digunakan metode pengujian *Chi-square*. *Dataset* yang digunakan dalam penelitian ini adalah Data Pokok Pendidikan di Kota Tegal.

#### 3.1. Prosedur Penelitian

Penelitian ini dilakukan dalam beberapa tahapan, yaitu tahapan identifikasi masalah, studi pustaka, pengumpulan data, pemodelan sistem, perancangan sistem, dan terakhir adalah implementasi sistem. Tahapan tersebut ditunjukkan pada Gambar 2.

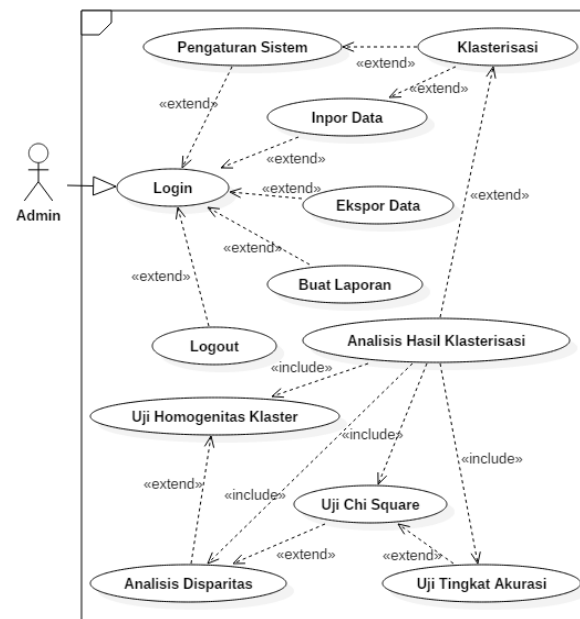


Gambar 2. Prosedur Penelitian

Tahapan-tahapan dalam Gambar 2 merupakan tahapan prosedur penelitian secara singkat, adapun tahapan prosedur secara lengkap dijelaskan sebagai berikut: Tahapan pertama merupakan tahapan identifikasi masalah. Pada tahapan ini dilakukan pengidentifikasian permasalahan disparitas tenaga pendidik sekolah dasar dan menengah yang terdapat di wilayah Kota Tegal. Tahapan kedua adalah tahapan studi pustaka. Pada tahapan ini dilakukan studi literatur terkait implementasi metode ilmiah dan

jenis pengujian yang digunakan, baik berupa jurnal, buku, maupun artikel ilmiah. Tahapan ketiga merupakan tahapan pengumpulan data. Pada tahapan ini dilakukan pengumpulan data melalui beberapa sumber data yang relevan. Dalam hal ini adalah data jumlah siswa dan ketersediaan tenaga pendidik (guru) sekolah dasar dan sekolah menengah di wilayah Kota Tegal yang dapat diperoleh dari dapodik (Data Pokok Pendidikan) wilayah Jawa Tengah. Selain itu, dalam penelitian ini juga menggunakan data rasio kebutuhan Guru sesuai dengan angka ideal rasio siswa-guru yang ada dalam Peraturan Pemerintah Nomor 74 tahun 2008 tentang Guru.

Keempat merupakan tahapan pemodelan sistem. Pada tahapan ini dilakukan pemodelan sistem menggunakan teknik pemodelan berorientasi objek yaitu *Unified Modeling Language (UML)*. Pemodelan disajikan dalam bentuk diagram *use case* (ditunjukkan pada Gambar 3) untuk menggambarkan kebutuhan fungsional perangkat lunak dan interaksinya dengan sistem eksternal dan penggunaanya, *activity diagram* untuk menggambarkan proses bisnis yang ada pada sistem, diagram alir atau flowchart untuk menjelaskan aktivitas secara teknis dan komperhensif sehingga dapat membantu dalam kontruksi sistem, dan *class diagram* untuk menunjukkan kelas yang ada dari sebuah sistem dan hubungannya secara logika serta menggambarkan struktur statis dan relasi antar kelas dari sebuah sistem.



Gambar 3. Use Case Diagram

Deskripsi dari setiap *use case* pada Gambar 3 adalah sebagai berikut: *Use case* Login merupakan fungsi dalam aplikasi yang bertujuan untuk keperluan autentikasi akun, untuk dapat menggunakan sistem,

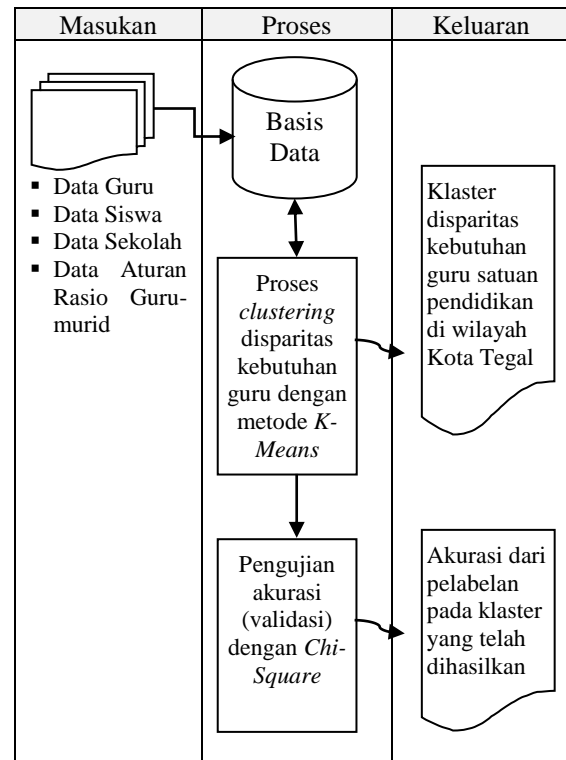
maka admin harus melakukan login terlebih dahulu ke dalam sistem. Setelah admin masuk ke *dashboard* sistem, admin dapat melakukan impor data. Admin harus melakukan impor data satuan pendidikan sebelum melakukan klusterisasi (jika sebelumnya data telah diimpor, maka tahapan ini bersifat opsional). Setelah data diimpor, admin dapat melakukan pengaturan sistem, pengaturan ini meliputi pengaturan default dan pengaturan data ideal rasio guru-murid pada satuan pendidikan. Fungsi pengaturan sistem tersebut digunakan untuk mengatur elemen-elemen yang dibutuhkan sistem agar proses klusterisasi dapat dilakukan. Selanjutnya admin dapat melakukan proses klusterisasi (*clustering*) terhadap data satuan pendidikan yang telah diimpor. Hasil dari fungsi ini berupa daftar satuan pendidikan sesuai dengan klaster yang dihasilkan.

Setelah hasil klaster didapatkan, admin dapat menggunakan fungsi Analisis Hasil Klusterisasi. Melalui fungsi tersebut, admin dapat melakukan analisis terhadap hasil klaster. Analisis ini digunakan untuk memberikan label pada setiap satuan pendidikan berdasarkan kondisi disparitasnya dan sekaligus mengetahui seberapa akurat pelabelan yang telah dilakukan. Terdapat 4 (empat) sub-fungsi dalam *use case* ini, yaitu: uji homogenitas klaster, analisis ketersediaan dan kebutuhan, uji chi square, dan validasi (uji tingkat akurasi). Setelah proses klusterisasi dan analisis selesai dilakukan, admin dapat menggunakan fungsi ekspor data. Fungsi ekspor data digunakan untuk melakukan ekspor data, baik data master maupun data hasil klusterisasi dalam bentuk tabel atau grafik. Selain fungsi ekspor data, admin juga dapat melakukan pembuatan laporan melalui fungsi Buat Laporan, admin dapat membuat laporan hasil klusterisasi dan kondisi disparitas ketersediaan guru. Fungsi terakhir dalam aplikasi ini adalah fungsi Logout, fungsi ini digunakan untuk mengakhiri sesi dan keluar dari sistem.

Tahapan kelima dari prosedur penelitian adalah Tahapan Perancangan Sistem. Pada tahapan ini dilakukan analisa terhadap tipe data dan perancangan basis data, pembentukan kerangka sistem informasi (ditunjukkan pada Gambar 4), dan perancangan antarmuka (*interfaace*) sesuai dengan perancangan basis data. Tahapan terakhir dalam prosedur penelitian adalah tahapan implementasi sistem.

Pada tahapan ini dilakukan pembangunan sistem berdasarkan pemodelan dan perancangan sistem yang telah dilakukan sebelumnya, dan mengimplementasikan metode *K-Means* dan *Chi-Square* dengan menggunakan bahan dan alat yang telah disiapkan. Selanjutnya, dilakukan analisis dan pengujian terhadap sistem yang telah dibangun untuk mengetahui apakah sistem sudah berjalan dengan baik dan sesuai dengan kebutuhan. Jika masih ditemukan ketidaksesuaian, maka perlu dilakukan perbaikan dan pengujian ulang sampai

sistem berjalan secara efektif sesuai dengan kebutuhan.



Gambar 4. Kerangka Sistem

Masukan pada Gambar 4 berupa data-data yang diperlukan oleh sistem yang meliputi data jumlah guru, jumlah siswa, data sekolah atau satuan pendidikan, dan data aturan (meliputi angka ideal rasio guru-murid berdasarkan peraturan pemerintah). Data-data tersebut akan disimpan oleh sistem ke dalam basis data. Selanjutnya data yang telah tersimpan tersebut dapat diproses oleh sistem untuk mengelompokkan satuan pendidikan berdasarkan disparitas kebutuhan guru menggunakan metode *K-Means*. Setelah hasil pengelompokan data didapatkan, selanjutnya dapat dilakukan pengujian akurasi dengan menggunakan *Chi-Square* sebagai validasi seberapa akurat pelabelan disparitas yang dihasilkan. Luaran dari sistem ini berupa klaster satuan pendidikan berdasarkan disparitas guru dan nilai akurasi hasil pelabelan dengan *Chi-square*.

### 3.2. Bahan dan Alat Penelitian

Data yang dibutuhkan dalam penelitian ini adalah data sekolah dasar dan menengah di wilayah Kota Tegal, dan data siswa dan guru di sekolah tingkat dasar dan menengah di wilayah Kota Tegal. Alat penelitian yang digunakan dalam penelitian ini terdiri dari 2 (dua) aspek, yaitu perangkat keras (meliputi komputer dengan spesifikasi minimum prosesor berfrekuensi 1 GHz dan RAM 4 GB), dan perangkat lunak (meliputi Netbeans IDE 8.2, JDK 1.8.0, StartUML 2.6.0 dan MySQL Server).

## 4. Hasil dan Pembahasan

### 4.1. Hasil Penelitian

Hasil dari penelitian ini berupa perangkat lunak Sistem Pendukung Keputusan (SPK) yang mengimplementasikan metode *K-Means Clustering* berbasis *Chi-square*. Perangkat lunak tersebut dapat menginformasikan disparitas kebutuhan guru pada setiap satuan pendidikan dan dapat melakukan klusterisasi berdasarkan disparitas yang dihasilkan, serta dapat menampilkan tingkat akurasi dari pelabelan pada kluster yang dihasilkan. Sehingga, aplikasi hasil penelitian ini dapat digunakan sebagai pendukung keputusan dalam mengidentifikasi profil ketersediaan Guru di wilayah Kota Tegal. Untuk memastikan apakah sistem yang telah dibangun telah sesuai dengan fungsi-fungsi yang didefinisikan pada tahapan analisis dan perancangan sistem, maka dilakukan pengujian sistem dengan menggunakan metode pengujian *blackbox*.

Hasil pengujian *blackbox* menunjukkan bahwa sistem dapat berjalan dengan baik sesuai dengan fungsionalitas yang dibutuhkan. Adapun untuk melakukan klusterisasi menggunakan aplikasi ini dapat dimulai dengan memasukan data dengan cara impor *file* pokok pendidikan ke dalam sistem (*file* berekstensi \*.xls atau \*.xlsx dan harus sesuai dengan format atau *template* dokumen yang telah disediakan). Selanjutnya, pengguna dapat melihat hasil kluster dengan memilih menu "kluster", secara otomatis sistem melakukan *clustering* dan selanjutnya pengguna dapat melakukan analisis terhadap kluster yang dihasilkan serta dapat melakukan pengujian akurasi melalui menu "analisis cluster" (Tabel 1).

Tabel 1. Sekolah dengan label disparitas tinggi

No.	Nama Satuan Pendidikan
1	SMP Atmaja Wacana
2	SMKN 3 Tegal
3	SMAS Muhammadiyah
4	SMAS Pancasakti Tegal
5	SMKS Muhammadiyah 1 Kota Tegal

Hasil pengelompokan (*clustering*) berupa 9 (sembilan) kluster, dan terdapat 3 (tiga) kluster dengan kondisi disparitas tinggi atau dengan kata lain ketersediaan Guru pada kluster tersebut masih kurang. Adapun status disparitas guru pada masing-masing kluster ditunjukkan pada Tabel 2, dan data sekolah dengan kondisi disparitas ketersediaan dalam kategori WAJAR. Sedangkan data spesifik satuan pendidikan yang memiliki label kondisi disparitas ketersediaan dalam kategori TINGGI dapat dilihat pada Tabel 2.

Tabel 2. Label umum masing-masing *cluster*

Cluster	Chi-Square	Label
1	20.43	Tinggi
2	10.08	Tinggi
3	2.557	Wajar
4	1.403	Wajar
5	3.446	Wajar
6	0.055	Wajar
7	16.23	Tinggi
8	0.437	Wajar
9	2.858	Wajar

Berdasarkan Tabel 1, maka terdapat 5 (lima) sekolah di wilayah Kota Tegal yang perlu mendapat perhatian lebih. Hal ini dikarenakan sekolah tersebut memiliki angka disparitas yang tinggi (terdapat perbedaan signifikan antara kebutuhan normatif dengan ketersediaan).

### 4.2. Pembahasan

Dalam proses menghasilkan kluster disparitas kebutuhan guru, sebelum dilakukan pengelompokan terlebih dahulu dilakukan *pre-processing* data. Proses ini bertujuan untuk memetakan nilai dari masing-masing variabel ke dalam rentang yang sama yaitu [0,1]. Sehingga pada saat perhitungan, proses *similarity* masing-masing variabel memberikan tingkat kepentingan yang sama (memberikan pengaruh yang sama). Setelah data dinormalisasi, berikutnya dilakukan klusterisasi dengan metode K-Means, dimana proses inialisasi centroid awal dilakukan secara random. Sedangkan penentuan jumlah kluster (biasa disebut K) dilakukan dengan menggunakan pendekatan *rule-of-thumbs*  $K = \sqrt{n/2}$ . Jumlah data yang digunakan dalam penelitian adalah 161 satuan pendidikan, sehingga jumlah  $K = \sqrt{161/2} = 8,9$  atau bisa dibulatkan menjadi  $K=9$ . Setelah klustering didapatkan, selanjutnya dilakukan uji homogenitas kluster untuk mengetahui struktur kluster. Pengujian homogenitas kluster dilakukan menggunakan *silhouette coefficient*. Adapun nilai rata-rata *silhouette* untuk seluruh nilai *silhouette* adalah 0.514 yang berarti hasil *clustering* memiliki struktur medium (baik).

Pada tahap selanjutnya, dilakukan pengujian dengan *chi-square* untuk memberikan label kepada masing-masing kluster pada umumnya atau masing-masing satuan pendidikan mengenai tingkat disparitas (perbedaan antara nilai ketersediaan dan nilai kebutuhan) guru pada satuan pendidikan di wilayah Kota Tegal. Karena digunakan 2 (dua) kategori, yaitu ketersediaan dan kebutuhan, maka nilai derajat kebebasan adalah  $(2-1) = 1$ . Berdasarkan nilai derajat kebebasan 1 dan toleransi kesalahan 0.5, maka nilai *Chi-square* adalah 3.841, sehingga berlaku aturan dalam pemberian label yaitu: hasil pelabelan

dianggap wajar jika nilai *chi-square* kurang dari 3.841, dan selain itu hasil pelabelan dianggap tinggi. Selanjutnya, dari hasil pengelompokan dan pengujian dilakukan evaluasi untuk mengukur tingkat akurasi atas pemberian label berdasarkan jumlah label terprediksi benar dibagi dengan jumlah keseluruhan label. Hasil evaluasi menunjukkan bahwa pelabelan memiliki tingkat akurasi sebesar 84.47% (akurat).

## 5. Kesimpulan

Implementasi metode *K-Means* berbasis *Chi-Square* berhasil diterapkan pada Sistem Pendukung Keputusan. Hasil uji *chi-square* menunjukkan bahwa dari 9 (sembilan) klaster yang dihasilkan, terdapat 3 (tiga) klaster satuan pendidikan yang memiliki angka disparitas TINGGI (terdapat perbedaan signifikan antara kebutuhan normatif dengan ketersediaan). Sedangkan hasil evaluasi menunjukkan bahwa tingkat akurasi pelabelan disparitas guru adalah 84.47%. Berdasarkan pemberian label tersebut, maka dapat diketahui bahwa sekolah dengan status disparitas TINGGI adalah SMP Atmaja Wacana, SMKN 3 Tegal, SMAS Muhammadiyah, SMAS Pancasakti Tegal, dan SMKS Muhammadiyah 1 Kota Tegal. Dengan demikian, sistem tersebut dapat digunakan sebagai media untuk mengidentifikasi profil ketersediaan Guru untuk masing-masing satuan pendidikan di Wilayah Kota Tegal dalam rangka pemenuhan ketersediaan Guru yang memadai.

Saran pada penelitian berikutnya adalah pengembangan pada aplikasi diharapkan dapat dikembangkan agar lebih dinamis (untuk seluruh wilayah di Indonesia) dan dapat diimplementasikan ke dalam aplikasi berbasis web dan mobile agar terintegrasi dengan website resmi Dinas Pendidikan di masing-masing wilayah di Indonesia.

## Ucapan Terima Kasih

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dalam penelitian ini, terutama kepada Politeknik Harapan Bersama yang telah memberikan pendanaan dalam penelitian ini.

## Daftar Pustaka

- Kemendikbud, 2016. Analisis Sumber Daya Manusia Pendidikan Dasar dan Menengah 2015/2016, Pusat Data dan Statistik Pendidikan dan Kebudayaan, Jakarta.
- Kuswantoro, E., Suprpto, K. Y., 2015. Penerapan algoritma k-means dengan optimasi jumlah cluster untuk pengelompokan angkatan kerja propinsi Jatim. *JAVA Journal of Electrical and Electronics Engineering* 3 (1), 58-62.
- Larose, T.D., 2014. *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley-Interscience, Canada.
- Manning, D.C., 2008. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge.
- Oyelade, J., Oladipupo, O., and Obagbuwa, C.I., 2010. Application of k means clustering algorithm for prediction of students academic performance. *International Journal of Computer Science and Information Security* (7), 292–295.
- Prasetyo, E., 2014. *Data Mining, Mengolah Data Menjadi Informasi Menggunakan Matlab*. ANDI, Yogyakarta.
- Stine, R., 2011. *Statistics for Business Decision Making and Analysis with Chi-Square Tests*. Addison-Wesley, Boston.
- Sunuyeko, N., Lani, A., and Wahyuni, L., 2017. Analisis kebutuhan guru dalam pengimplementasian kurikulum 2013 di Sekolah Dasar. *Portal Jurnal Elektronik Universitas Negeri Malang* 2, 1-9.
- Wu, X., Wu, B., Sun, J., Qiu, S., and Li, X., 2014. A Hybrid fuzzy k-harmonic means clustering algorithm. *Applied Mathematical Modelling* 39(12), 3398-3409.
- Wu, X., Kumar, V., Quinlan, R.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, J.G., Ng, A., Liu, B., Yu, S.P., Zhou, H., Steinbach, M., Hand, J.D., and Steinberg, D., 2008. Top 10 algorithms in data mining, *Knowledge and Information Systems* 14 (1), 1-37.