

# PENERAPAN ALGORITMA K-MEANS DALAM PENENTUAN JURUSAN SISWA SMAN 05 BOMBANA

Muh. Sulkifly Said, Yusti  
STMIK CATUR SAKTI KENDARI,  
Jalan. Drs. H. Abdullah Silondae No. 109 Kendari  
kiflinux@gmail.com, yusti.si35@gmail.com

SMA 05 Bombana, merupakan salah satu sekolah di Kel. Dongkala Kec. Kabaena Timur Kab. Bombana Sulawesi Tenggara yang mengadakan penjurusan siswa ke dalam 3 jurusan yaitu IPA, IPS dan Bahasa. Kendala yang sering ditemukan dalam proses penjurusan di SMAN 05 Bombana yaitu sulitnya menentukan siswa mana yang memenuhi kriteria untuk menempati jurusan tertentu. Hal ini dikarenakan proses penjurusan masih dilakukan dengan manual, sehingga membutuhkan waktu yang cukup lama dan dinilai kurang tepat. Pada penelitian ini akan diangkat suatu kasus yaitu menganalisa algoritma K-Means untuk sistem pendukung keputusan penjurusan siswa di SMAN 05 Bombana. Algoritma K-Means dapat mengelompokkan (segmentasi) data yang mempunyai atribut dan mempunyai jumlah data yang banyak, sehingga dapat dimanfaatkan dalam sistem penentuan penjurusan siswa yang sesuai dengan kemampuan akademik siswa. Berdasarkan hasil implementasi terhadap algoritma K-Means untuk sistem pendukung keputusan penjurusan, maka kesimpulan yang dapat diambil adalah algoritma K-Means dapat diterapkan dalam sistem pendukung keputusan penjurusan siswa

**Kata kunci :** SPK, Jurusan, K-Means, Nilai SMAN 05 Bombana

## I. PENDAHULUAN

Perkembangan teknologi dewasa ini semakin pesat seiring perkembangan zaman. Dengan perkembangan teknologi yang semakin canggih, komunikasi dapat dilakukan dengan berbagai cara. Perkembangan teknologi dapat digunakan untuk mengolah data, termasuk memproses, mendapatkan, menyusun, menyimpan, memanipulasi data dalam berbagai cara untuk menghasilkan informasi yang berkualitas, yaitu informasi yang relevan, akurat dan tepat waktu, yang digunakan untuk keperluan pribadi, bisnis, dan pemerintahan.

Kemajuan teknologi informasi telah menyebabkan banyak orang dapat memperoleh data dengan mudah bahkan cenderung berlebihan. Data tersebut semakin lama semakin banyak data terakumulasi, akibatnya pemanfaatan data yang terakumulasi tersebut menjadi

tidak optimal. Sehingga salah satu kebutuhan dari organisasi swasta maupun pemerintah dalam menyediakan sarana informasi adalah informasi yang berbasis teknologi. Dimana sistem yang telah menggunakan komputerisasi memiliki tingkat keakuratan yang jauh lebih efektif dan efisien dalam pengolahan data bila dibandingkan dengan sistem manual, sehingga kebutuhan akan adanya komputerisasi dalam pengolahan data sangatlah dibutuhkan.

SMAN 05 Bombana yang berdiri sejak tahun 2004 merupakan salah satu sekolah menengah atas di Kabupaten Bombana yang terletak di jalan pendidikan no. 3 Dongkala, Kab. Bombana Sulawesi Tenggara. Saat ini SMA tersebut telah memiliki tiga jurusan yaitu IPA, IPS dan bahasa. Peminatan atau lebih dikenal dengan penjurusan bertujuan untuk memfokuskan atau mengarahkan materi pembelajaran siswa-siswi sekolah menengah agar sesuai dengan nilai akademik yang dimiliki oleh para siswa-siswi di sekolah.

Hal ini sejalan dengan Permendikbud 81 A tahun 2013 tentang implementasi kurikulum 2013 menjelaskan bahwa pendidikan merupakan proses yang sistematis manusia secara holistik yang memungkinkan potensi diri berkembang secara optimal. Peminatan atau penjurusan ini juga membantu memberi gambaran akan bidang yang nantinya ditekuni oleh para siswa-siswi setelah lulus sekolah nanti. Namun dalam penentuan jurusan untuk SMA sering sekali menjadi permasalahan di sekolah, antara lain sulitnya menentukan siswa mana yang masuk kriteria untuk menempati jurusan tertentu. Kriteria penentuan jurusan siswa hanya berdasarkan nilai akademik, penentuan jurusan dengan sistem yang lama ditentukan berdasarkan nilai akhir saja maksudnya bahwa penentuan hanya berdasarkan nilai rata-rata USBN (Ujian Sekolah Berstandar Nasional) tanpa menjabarkan setiap nilai dari mata pelajaran yang mewakili jurusan tersebut.

Dengan memanfaatkan algoritma *K-Means* data yang ada akan dikelompokkan menjadi beberapa *cluster* (kelompok) yang masing-masing diwakili oleh pusat *cluster*. Data yang ada diklasifikasikan berdasarkan hasil akhir evaluasi nilai akademik yang diambil sejak pendaftaran serta menjabarkan matapelajaran yang mewakili setiap jurusan yang ada. Sehingga dengan adanya sistem ini membantu guru dalam menentukan jurusan yang sesuai dengan nilai akademik yang telah dijadikan syarat dalam penentuan jurusan dan algoritma

*K-Means* ini adalah algoritma yang hanya mempartisi data kedalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain.

Dalam jurnal (Manurung, 2014) yang berjudul “Perancangan Aplikasi Penentuan Jurusan di SMA Menggunakan *K-Means*”, data yang digunakan untuk proses analisis *K-Means* adalah data nilai yang kemudian dilakukan *clustering* dengan dua cluster/kelompok yaitu jurusan IPA dan IPS. Dari jurnal tersebut ingin dikembangkan dengan menambahkan cluster yang lain yaitu cluster bahasa karena jurusan yang ada sekarang ini sudah mengalami penambahan jurusan yaitu jurusan bahasa.

Berdasarkan permasalahan tersebut, maka dalam penelitian ini dirancang sebuah sistem yang dituangkan dalam karya ilmiah yang berjudul “*Penerapan Algoritma K-Means dalam Penentuan Jurusan Siswa SMA Negeri 05 Bombana*”

## II. LANDASAN TEORI

### 2.1 Sistem Informasi

#### A. Pengertian Sistem

*Sistem berasal dari bahasa latin (system) dan bahasa Yunani (Sustema) adalah suatu kesatuan yang terdiri dari komponen atau elemen yang dihubungkan bersama untuk memudahkan aliran informasi, materi atau energi. Istilah ini sering digunakan untuk menggambarkan suatu set entitas yang berinteraksi, dimana suatu model matematika sering dibuat (Husda, 2012).*

#### B. Pengertian Informasi

*Menurut Husda (2012). Secara umum informasi dapat didefinisikan sebagai hasil dari pengolahan data dalam suatu bentuk yang lebih berguna dan lebih berarti bagi penerimanya yang menggambarkan suatu kejadian-kejadian yang nyata yang digunakan untuk pengambilan keputusan.*

#### C. Pengertian Sistem Informasi

Menurut Prahasta (2005) bahwa sistem informasi adalah suatu *entity* (kesatuan) formal yang terdiri dari berbagai sumber daya fisik maupun logika. Defenisi lain menyatakan bahwa sistem informasi adalah suatu sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian yang mendukung fungsi operasi organisasi yang bersifat manajerial dengan kegiatan strategi dari suatu organisasi untuk dapat menyediakan laporan-laporan yang diperlukan oleh pihak luar tertentu (Sutabri, 2012).

#### D. Sasaran Sistem Informasi

- 1) Dengan adanya sistem informasi maka penyelesaian tugas atau pekerjaan akan semakin meningkat.
- 2) Proses pengerjaan tugas atau pekerjaan akan mempunyai nilai efektivitas yang tinggi secara keseluruhan.

- 3) User dituntut untuk lebih produktif supaya memperoleh output yang berkualitas.
- 4) Output yang berkualitas akan mendatangkan pendapatan atau keuntungan yang lebih besar dari pada biaya pembuatan dan maintenance (perawatan) sistem itu sendiri.

### 2.2 Data

Berikut literatur tentang data “Data” oleh (Kusrini, 2007). Informasi sangat erat kaitannya dengan data, karena informasi berasal dari data. Data inilah yang akan dipilih, dimodifikasi atau diperbaharui oleh program-program agar menjadi sebuah informasi baru yang dapat bermanfaat. Data dapat dianalogikan bahan mentah, yang harus diolah untuk menjadi barang jadi yang siap pakai. Data merupakan kata jamak dari datum yang berarti gambaran mengenai fakta, statistik dan lain sebagainya.

### 2.3 Pendekatan Penjurusan

Sesuai dengan pasal 12 ayat 2 Undang-undang no 20 tahun 2003 tentang sistem pendidikan yang menyatakan bahwa setiap peserta didik berhak mendapatkan pelayanan sesuai dengan bakat, minat dan kemampuannya. Pasal ini menjelaskan bahwa setiap siswa sekolah, khususnya sekolah menengah atas ataupun sederajat mempunyai hak dalam menentukan pilihan pembelajaran sesuai dengan potensi, minat dan bakat yang dimiliki oleh para siswa-siswi. Pihak sekolah wajib menyediakan jurusan atau peminatan agar siswa-siswi dapat melanjutkan pengembangan yang dimiliki oleh masing-masing individu.

### 2.4 K-Means

#### 2.4.1 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma pengelompokan *iterative* yang melakukan partisi set data kedalam sejumlah *cluster* yang sudah ditetapkan awal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, *relative* cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi salah satu algoritma yang sering di pakai dalam bidang *data mining* (Wu dan Kumar, 2009).

*Cluster* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan clusteran. Objek yang didalam *cluster* memiliki karakteristik satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering*. Oleh karena itu, *clustering* atau *cluster* sangat berguna dan bisa menemukan *group* atau kelompok yang tidak dikenal dalam data. *Clustering* banyak digunakan dalam berbagai aplikasi seperti misalnya pada *business intelegence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). Didalam *business intelegence*, *clustering* bisa mengatur banyak customer ke dalam banyaknya kelompok. Contohnya mengelompokkan *customer* ke dalam beberapa *cluster* dengan kesamaan karakteristik yang kuat. *Clustering* juga di kenal sebagai data segmentasi karena *clustering* mempartisi banyak data

set ke dalam banyak *group* berdasarkan kesamaannya. Selain itu *clustering* bisa sebagai *outlier detection*.

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. Pada awal iterasi pusat setiap *cluster* ditetapkan secara bebas (sembarang),  $C_{kj}$  ( $k=1, \dots, k; j=1, \dots, m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan perhitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean, seperti pada persamaan (2.1), yaitu :

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

Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak data tersebut ke pusat *cluster* ke- $k$  bernilai paling kecil jika dibandingkan dengan jarak kepusat *cluster* lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2.2) selanjutnya, kelompokan data-data yang menjadi anggota pada setiap *cluster*.

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - C_{ij})^2}$$

Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut, dengan menggunakan rumus pada persamaan (2.3) :

$$C_{ij} = \frac{\sum_{i=1}^p x_{ij}}{p}$$

Dimana  $x_{ij} \in \text{cluster ke } k$   
 $p = \text{banyaknya anggota cluster ke } k$

**2.4.2 Manfaat Cluster**

*Cluster* merupakan metode segmentasi data yang sangat berguna dalam prediksi dan analisa masalah bisnis tertentu (Berson dan Smith, 2001). Misalnya segmentasi pasar, marketing dan pemetaan zonasi wilayah. Serta identifikasi objek dalam berbagai bidang seperti *computer vision* dan *image processing*.

**2.4.3 Langkah-Langkah K-Means**

Proses *clustering* dengan menggunakan algoritma *K-Means* memiliki langkah-langkah sebagai berikut :

- a. Tentukan jumlah *cluster* yang diinginkan jika perlu tetapkan ambang batas perubahan *centroid*.
- b. Tentukan titik pusat *cluster*
- c. Hitung jarak objek dengan *centroid*
- d. Kelompokkan objek berdasarkan jarak terdekat
- e. Ulangi langkah dari awal sampai kelompok data yang dihasilkan dari perhitungan sudah mencapai ambang batas yang ditetapkan.

**2.5 Bagan Alir(Flowchart)**

Diagram alir atau *flowchart* merupakan sebuah diagram yang digambarkan secara grafis yang menyatakan aliran proses atau algoritma yang menampilkan urutan-

urutan atau langkah-langkah dari proses atau algoritma itu sendiri (Nugroho, 2005).

**1. Jenis-jenis bagan alir/Flowchart**

Ada beberapa jenis - jenis flowchart diantaranya:

- a) Bagan Alir Sistem (*System Flowchart*)
- b) Bagan Alir Dokumen (*Document Flowchart*)
- c) Bagan Alir Skematik (*Schematic Flowchart*)
- d) Bagan Alir Program (*Program Flowchart*)
- e) Bagan Alir Proses (*Process Flowchart*)

**2. Simbol-simbol flowchart**

Berikut beberapa simbol yang dalam *flowchart* yang dapat dilihat pada gambar 2.1 berikut.

| NO | Simbol | Keterangan                | Maksud   |
|----|--------|---------------------------|--|
| 1  |        | Terminal                  | Menyatakan awal dan akhir proses   |
| 2  |        | Input/output              | Untuk memasukkan data dari luar atau hasil dari suatu proses                 |
| 3  |        | Proses                    | Menyatakan proses yang dilakukan   |
| 4  |        | Flowline                  | Menunjukkan alur proses yang dilakukan                                       |
| 5  |        | Decision                  | Untuk menentukan keputusan dari sebuah pertanyaan                            |
| 6  |        | Output "hard copy"        | Menentukan hasil dari format yang dicetak                                    |
| 7  |        | On page connector         | Penghubung bagian flowchart pada satu halaman                                |
| 8  |        | Off page connector        | Penghubung flowchart pada halaman yang berbeda                               |
| 9  |        | On-line storage           | Untuk menggunakan penyimpanan akses langsung                                 |
| 10 |        | Input manual              | Memasukkan data melalui papan ketik  |
| 11 |        | Preparation               | Menyatakan persiapan awal dari proses yang dilakukan                         |
| 12 |        | Predefined process        | Permulaan sub program / proses menjalankan sub program                       |
| 13 |        | Symbol magnetig-tape unit | Menyatakan input berasal pita magnetic atau output disimpan ke pita magnetic |

Gambar 2.1 Simbol-simbol *Flowchart*

**2.6 Sistem Basis Data**

*Database* adalah suatu susunan/kumpulan data operasional lengkap dari suatu organisasi/perusahaan yang diorganisi/kelola dan simpan secara terintegrasi dengan menggunakan metode tertentu dengan menggunakan komputer sehingga mampu menyediakan informasi yang diperlukan pemakainya. Distem *database* adalah suatu sistem penyusunan dan pengelolaan *record-record* dengan menggunakan komputer. Dengan tujuan untuk menyimpan atau merekam serta memelihara data operasional lengkap sebuah organisasi/perusahaan sehingga mampu menyediakan informasi yang diperlukan pemakai untuk kepentingan pengambilan keputusan.

**2.7 Sistem Development Life Cycle**

SDLC (*Sistem Development Life Cycle*) merupakan pengembangan yang berfungsi sebagai sebuah mekanisme untuk mengidentifikasi perangkat lunak. Pengembangan sistem informasi yang berbasis komputer dapat merupakan tugas kompleks yang membutuhkan

banyak sumber daya dan dapat memakan waktu untuk menyelesaikannya.

*System Development Life Cycle* (SDLC) menurut Kendall (2011) dalam bukunya mendefinisikan SDLC sebagai suatu pendekatan yang memiliki tahapan atau bertahap untuk melakukan analisa dan membangun suatu sistem dengan menggunakan siklus yang lebih spesifik terhadap kegiatan pengguna yang secara garis besar terbagi menjadi tiga kegiatan utama, yaitu :

1. *Analysis*,
2. *Design*,
3. *Implementasion*

## 2.8 Perbedaan MySql dan SQL

MySQL dan SQL adalah dua “mahluk” yang berbeda. Mudahnya MySQL adalah softwarenya sedangkan SQL adalah bahasa perintahnya. Awalnya MySQL dijalankan pada sistem operasi Unix dan Linux. Tapi para penggemar windows pun sekarang sudah tersedia MySQL versi mereka yang menggunakan linux (*RedHat*, *Mandrake*) biasanya SQL sudah terinstal secara *default*. Bila belum bisa diinstall maka dengan mudah dapat menginstalnya menggunakan RPM (*Redhat Package Manage*), merupakan prosedur instalasi *software* yang menjadi standar de-facto untuk linux, walaupun anda tidak menggunakan linux *RedHat*.

MySQL merupakan *database* yang paling digemari sebagai piranti perangkat lunak yang *open source*, dengan alasan program ini merupakan *database* yang sangat kuat dan cukup stabil untuk digunakan sebagai media penyimpanan data. Sebagai sebuah *database server* yang mampu untuk manajemen *database* dengan baik, MySQL terhitung merupakan *database* yang paling digemari dan paling banyak digunakan dibanding *database* lainnya (Nugroho, 2005).

## 2.9 PHP

*Hypertext Preprocessor* atau biasa disingkat dengan istilah *PHP* merupakan bahasa pemrograman *server side* yang paling populer dan yang paling banyak digunakan. Pada awalnya *PHP* memiliki singkatan *Personal Home Page Tool* yang pertama kali dibuat oleh Rasmus Lerdford. Namun sekarang *PHP* sudah dibuat dan dilengkapi oleh banyak pihak sehingga mengalami perkembangan yang sangat pesat. Singkatannya juga telah berubah menjadi singkatan berulang, yakni *PHP Hypertext Preprocessor*.

Dengan adanya *PHP*, dunia situs *website* menjadi lebih menarik dan interaktif karna dengan menggunakan *PHP*, para pengunjung bisa saling berkomunikasi satu sama lain.

Di bawah ini beberapa keunggulan yang dimiliki oleh *PHP Hypertext Preprocessor* (Ariyanto, 2005) :

- a. Mendukung penggunaan *cookie*.
- b. Memiliki fitur dalam menangani *session*.
- c. Mendukung integrasi dengan *database*.
- d. Mendukung *regular expression*.

Penanganan kesalahan yang dibuat secara bertingkat.

## 2.10 XAMP

*XAMPP* adalah sebuah *software web server apache* yang didalamnya sudah tersedia *database server MySQL* dan dapat mendukung pemrograman *PHP*. *XAMPP* merupakan *software* yang mudah digunakan, gratis dan mendukung instalasi di *Linux* dan *Windows*. Keuntungan lainnya adalah cuma menginstal satu kali sudah tersedia *Apache Web Server*, *MySQL Database Server*, *PHP Support* (*PHP 4* dan *PHP 5*) dan beberapa modul lainnya. *XAMPP* adalah perangkat lunak bebas, yang mendukung banyak sistem operasi, merupakan kompilasi dari beberapa program. Fungsinya adalah sebagai *server* yang berdiri sendiri (*localhost*), yang terdiri atas program *Apache HTTP Server*, *MySQL database*, dan penerjemah bahasa yang ditulis dengan bahasa pemrograman *PHP* dan *Perl*. Nama *XAMPP* merupakan singkatan dari X (empat sistem operasi apapun), *Apache*, *MySQL*, *PHP* dan *Perl*. Program ini tersedia dalam GNU (*General Public License*) dan bebas, merupakan *web server* yang mudah digunakan yang dapat melayani tampilan halaman *website* yang dinamis.

## III. METODE PENELITIAN

### 3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada SMA Negeri 05 Bombana yang beralamat di jalan pendidikan no. 3 Kelurahan Dongkala Kecamatan Kabaena Timur Kabupaten Bombana Sulawesi Tenggara. Penelitian ini dilaksanakan mulai bulan Mei sampai Juni 2019.

### 3.2 Jenis dan Sumber Data

#### a. Jenis Data

Data yang digunakan dalam penelitian ini merupakan data kuantitatif yaitu data nilai siswa tahun ajaran 2018/2019 yang artinya penelitian yang dilakukan adalah menekankan analisisnya pada data-data *numeric* (angka), yang bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan berdasarkan data yang diperoleh dengan cara menyajikan, mengumpulkan dan menganalisis data tersebut sehingga menjadi informasi baru yang dapat digunakan untuk menganalisa mengenai masalah yang diteliti. Sedangkan sumber data yang diambil adalah data primer, yaitu data penelitian yang diperoleh langsung dari SMAN 05 Bombana.

#### b. Sumber Data

Data yang digunakan dalam penelitian ini bersumber dari SMA Negeri 05 Bombana khususnya pada bagian akademik sekolah. Sehingga berdasarkan sumber data tersebut disebut data primer.

### 3.3 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah :

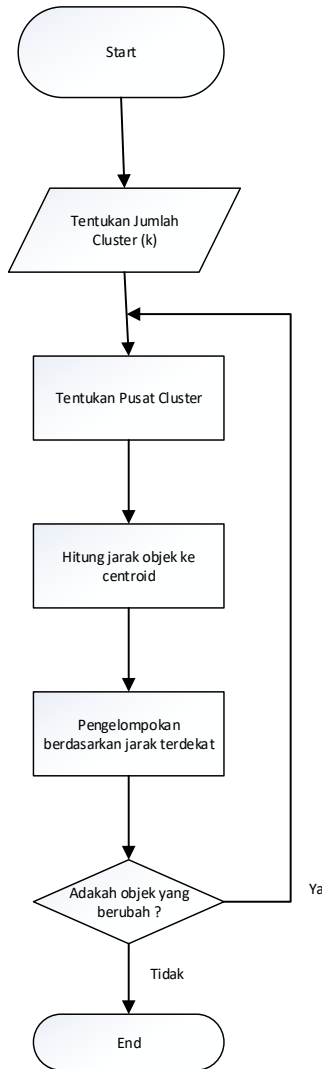
#### 3.3.1 Wawancara

Wawancara yaitu suatu metode pengumpulan data dengan memberikan pertanyaan-pertanyaan kepada pihak-pihak yang berkaitan dengan penelitian yang dapat memberikan informasi dan keterangan terkait penelitian dalam hal ini bagian Akademik pada SMA Negeri 05 Bombana.

3.3.2 Kepustakaan

Dilakukan dengan membaca berbagai literature, bahan-bahan terbitan yang mempunyai hubungan dengan penelitian ini.

3.4 Flowchart K-Means



Gambar 3.1 Flowchart K-means

3.5 Metode Pengembangan Sistem

Metode pengembangan sistem yang di pakai adalah metode SDLC (Sistem Development Life Cycle) yang terdiri dari tiga kegiatan secara garis besar yaitu :

3.5.1 Analysis

Proses penjurusan yang dijalankan di SMAN 05 Bombana masih dilakukan secara manual Bagian penjurusan yang menentukan siswa mana yang masuk kejurusan tertentu berdasarkan kriteria yang ada dilakukan berdasarkan nilai USBN(Ujian Sekolah Berstandar Nasional) SMP siswa serta peminatan pada saat pendaftaran siswa baru. Belum ada sistem yang mereka gunakan untuk proses penjurusan tersebut. Adapun sistem yang akan dibuat adalah sebuah aplikasi dengan menggunakan algoritma K-Means dalam

menentukan jurusan yang sesuai dengan nilai akademik siswa/i di SMAN 05 Bombana. Aplikasi yang akan dibuat membutuhkan data siswa dan data nilai siswa yang diperoleh dari pihak sekolah. Sistem yang akan dibuat bertujuan untuk memberikan informasi dalam menentukan jurusan siswa yang sesuai dengan kriteria yang telah ditentukan sekolah.

3.5.2 Design System

Apabila proses analisa telah selesai kemudian dilanjutkan dengan tahap *design system*, tahap ini membahas tentang bagaimana membuat *design system* pada sebuah aplikasi. Dalam *design* ini penulis akan membuat *flowchart program* dan tampilan *interface* aplikasi.

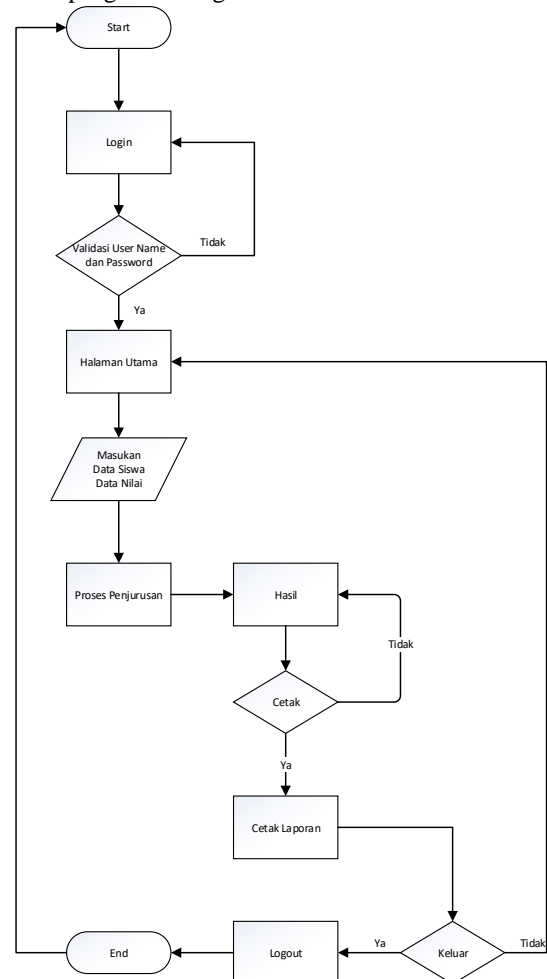
3.5.3 Implementation

Tahap selanjutnya adalah menulis kode program yang akan disesuaikan oleh komputer, dan melakukan penyesuaian perhitungan *k-means* yang dilakukan secara manual dengan aplikasi yang dibangun. Dalam tahap ini penulis membuat aplikasi program dengan bahasa pemrograman PHP dan menggunakan *MySQL* sebagai tempat penyimpanan *database*.

3.6 Perancangan Sistem

Pada perancangan sistem yang akan dibuat penulis merancang sistem dengan merancang *flowchart program*.

Flowchart program sebagai berikut :



Gambar 3.2 Flowchart Program

2.7 Perancangan *Interface*

Perancangan interface bertujuan untuk memberikan penjelasan terhadap Program yang akan dihasilkan. Perancangan interface akan memberikan gambaran dari setiap form-form dalam program. Perancangan interface tampak pada gambar-gambar berikut :

1. Desain Login

Gambar 3.3. Desain Login

2. Desain Menu Home

Gambar 3.4. Desain Menu Home

3. Desain Siswa

Gambar 3.5. Desain Data Siswa

4. Desain Data Nilai

Gambar 3.6. Desain Data Nilai

5. Desain Data Penjurusan

Gambar 3.7. Desain Data Penjurusan

6. Desain Laporan

Gambar 3.8. Desain Laporan



IV. HASIL DAN PEMBAHASAN

4.1. Hasil Pengolahan Data Secara Manual

. Untuk dapat melakukan penentuan jurusan data-data tersebut akan di hitung sehingga menjadi beberapa cluster(kelompok) dan perlu dilakukan beberapa langkah untuk proses perhitungan, yaitu :

Kriteria yang digunakan untuk melakukan perhitungan k-means ini ditentukan sesuai dengan kebutuhan analisis, dimana kriteria yang digunakan diambil dari nilai USBN (Ujian Sekolah Berstandar Nasional) SMP. Adapun daftar kriteria matapelajaran yang digunakan dapat dilihat pada tabel 3.1, yaitu :

Tabel 4.1 kriteria yang digunakan

| No | Nilai USBN       | Variabel |
|----|------------------|----------|
| 1  | IPA              | X1       |
| 2  | IPS              | X2       |
| 3  | Bahasa Inggris   | X3       |
| 4  | Bahasa Indonesia | X4       |
| 5  | Matematika       | X5       |

Contoh data awal sebelum dilakukan perhitungan untuk penentuan jurusan dengan matapelajaran yang menjadi kriteria penjurusan, dari masing-masing data dapat dilihat dari tabel 4.2 yaitu :

Tabel 4.2 Sampel Data Nilai Siswa yang Menjadi Kriteria

| NIS     | IPA  | IPS  | B.Indonesia | B.inggris | Mtk |
|---------|------|------|-------------|-----------|-----|
| NIS0001 | 8    | 8    | 7           | 8         | 6   |
| NIS0002 | 7    | 7,9  | 8           | 8         | 8   |
| NIS0003 | 9    | 8,3  | 9           | 9         | 8   |
| NIS0004 | 6,5  | 7    | 7           | 8,8       | 7   |
| NIS0005 | 7    | 7,9  | 7,8         | 7,7       | 8   |
| NIS0006 | 8,35 | 8,35 | 9           | 6,7       | 6   |

Sumber : Data Olahan Penulis

Selanjutnya akan digunakan algoritma k-means untuk melakukan pengelompokan data yang ada. Data yang ada di kelompokkan menjadi 3 kelompok. Adapun langkah perhitungannya sebagai berikut :

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi tiga *cluster* yaitu cluster ke-satu mewakili jurusan IPA, cluster ke-dua mewakili jurusan IPS dan cluster ke-tiga mewakili jurusan Bahasa ini diperoleh dari jumlah jurusan yang ada di SMAN 05 Bombana.
2. Tentukan titik pusat cluster awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal *cluster* ditentukan dengan memilih siswa yang mempunyai nilai yang mewakili setiap *cluster* dalam hal ini *cluster*-ke 1 mewakili jurusan IPA dengan siswa yang mempunyai nilai tertinggi, *cluster*-ke 2 mewakili jurusan IPS dengan siswa yang mempunyai nilai sedang dan *cluster*-ke 3 mewakili jurusan bahasa dengan siswa yang mempunyai nilai terendah. sehingga didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel 4.3.

Tabel 4.3 Pusat Awal Cluster

| Cluster | X1  | X2  | X3 | X4  | X5 |
|---------|-----|-----|----|-----|----|
| C1      | 9   | 8,3 | 9  | 9   | 8  |
| C2      | 8   | 8   | 7  | 8   | 6  |
| C3      | 6,5 | 7   | 7  | 8,8 | 7  |

Keterangan :

- X1 = Nilai IPA
- X2 = Nilai IPS
- X3 = Nilai Bahasa Indonesia
- X4 = Nilai Bahasa Inggris
- X5 = Nilai Matematika

3. Hitung jarak setiap data yang ada terhadap setiap pusat *cluster*. Misalkan untuk perhitungan jarak data siswa pertama dengan pusat cluster pertama dengan rumus euclidian distance yaitu :

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

Perhitungan jarak data pertama dengan pusat cluster pertama adalah :

$$d_{11} = \sqrt{(8-9)^2 + (8-8,3)^2 + (7-9)^2 + (8-9)^2 + (6-8)^2} = 3,176476$$

Perhitungan jarak data pertama dengan pusat cluster kedua adalah :

$$d_{12} = \sqrt{(8-8)^2 + (8-8)^2 + (7-7)^2 + (8-8)^2 + (6-6)^2} = 0$$

Perhitungan jarak data pertama dengan pusat cluster ketiga adalah :

$$d_{13} = \sqrt{(8-6,5)^2 + (8-7)^2 + (7-7)^2 + (8-8,8)^2 + (6-7)^2} = 2,211334$$

Tabel 4.4 Hasil Perhitungan Jarak Data

| NIS     | X1   | X2   | X3  | X4  | X5 | C1   | C2   | C3   |
|---------|------|------|-----|-----|----|------|------|------|
| NIS0001 | 8    | 8    | 7   | 8   | 6  | 3,17 | 0    | 2,21 |
| NIS0002 | 7    | 7,9  | 8   | 8   | 8  | 2,48 | 2,45 | 1,92 |
| NIS0003 | 9    | 8,3  | 9   | 9   | 8  | 0    | 3,17 | 3,60 |
| NIS0004 | 6,5  | 7    | 7   | 8,8 | 7  | 3,60 | 2,21 | 0    |
| NIS0005 | 7    | 7,9  | 7,8 | 7,7 | 8  | 2,7  | 2,39 | 1,97 |
| NIS0006 | 8,35 | 8,35 | 9   | 6,7 | 6  | 3,11 | 2,43 | 3,38 |

4. Setelah menghitung jarak data ke centroidnya, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimumnya. Suatu data akan menjadi anggota dari cluster yang memiliki jarak terdekat (terkecil) dari pusat *cluster*-nya.

Rumus yang digunakan adalah :

$$\text{Min } \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - C_{ij})^2}$$

Minimum dari  $d_{11} = (C1_1, C2_1, \text{ dan } C3_1)$   
 Minimum dari  $d_{11} = (3,176476) (0) (2,211334) = 0$   
 Minimum dari  $d_{12} = (C1_2, C2_2, \text{ dan } C3_2)$

Minimum dari  $d_{12} = (2,481935) (2,45153) (1,923538) = 1,923538$

Tabel 4.5 Data Jarak Minimum (Jarak Terpendek)

| NIS     | X1   | X2   | X3  | X4  | X5 | C1    | C2   | C3   | Jrk MIN |
|---------|------|------|-----|-----|----|-------|------|------|---------|
| NIS0001 | 8    | 8    | 7   | 8   | 6  | 3,176 | 0    | 2,21 | 0       |
| NIS0002 | 7    | 7,9  | 8   | 8   | 8  | 2,481 | 2,45 | 1,92 | 1,92    |
| NIS0003 | 9    | 8,3  | 9   | 9   | 8  | 0     | 3,17 | 3,60 | 0       |
| NIS0004 | 6,5  | 7    | 7   | 8,8 | 7  | 3,602 | 2,21 | 0    | 0       |
| NIS0005 | 7    | 7,9  | 7,8 | 7,7 | 8  | 2,7   | 2,39 | 1,97 | 1,97    |
| NIS0006 | 8,35 | 8,35 | 9   | 6,7 | 6  | 3,116 | 2,43 | 3,38 | 2,43    |

Tabel 4.6 Centroid Baru

| IPA  | IPS  | B.Indo | B.inggris | Mtk | C1 | C2 | C3 | Penjurusan |
|------|------|--------|-----------|-----|----|----|----|------------|
| 8    | 8    | 7      | 8         | 6   |    | *  |    | IPS        |
| 7    | 7,9  | 8      | 8         | 8   |    |    | *  | BAHASA     |
| 9    | 8,3  | 9      | 9         | 8   | *  |    |    | IPA        |
| 6,5  | 7    | 7      | 8,8       | 7   |    |    | *  | BAHASA     |
| 7    | 7,9  | 7,8    | 7,7       | 8   |    |    | *  | BAHASA     |
| 8,35 | 8,35 | 9      | 6,7       | 6   |    | *  |    | IPS        |

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terpendek antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat. Dengan cara membandingkan hasil cluster dan diambil yang paling kecil.

5. Berdasarkan pengelompokan tersebut, selanjutnya mencari centroid baru berdasarkan membership dari masing-masing cluster yaitu menghitung rata-rata dari data masing-masing cluster.

Rumus yang digunakan adalah :

$$C_{ij} = \frac{\sum_{i=1}^p X_{ij}}{p}$$

Perhitungan cluster pertama sebagai berikut :

$$C_1 = X1_3 / 1 = 9 / 1 = 9$$

$$C_1 = X2_3 / 1 = 8,3 / 1 = 8,3$$

$$C_1 = X3_3 / 1 = 9 / 1 = 9$$

$$C_1 = X4_3 / 1 = 9 / 1 = 9$$

$$C_1 = X5_3 / 1 = 8 / 1 = 8$$

Perhitungan cluster kedua sebagai berikut :

$$C_2 = (X1_1 + X1_6) / 2 = (8 + 8,35) / 2 = 8,175$$

$$C_2 = (X2_1 + X2_6) / 2 = (8 + 8,35) / 2 = 8,175$$

$$C_2 = (X3_1 + X3_6) / 2 = (7 + 9) / 2 = 8$$

$$C_2 = (X4_1 + X4_6) / 2 = (8 + 6,7) / 2 = 7,35$$

$$C_2 = (X5_1 + X5_6) / 2 = (6 + 6) / 2 = 6$$

Perhitungan cluster ketiga sebagai berikut :

$$C_3 = (X1_2 + X1_4 + X1_5) / 3 = (7 + 6,5 + 7) / 3 = 6,83333333$$

$$C_3 = (X1_2 + X2_4 + X2_5) / 3 = (7,9 + 7 + 7,9) / 3 = 7,6$$

$$C_3 = (X3_2 + X3_4 + X3_5) / 3 = (8 + 7 + 7,8) / 3 = 7,6$$

$$C_3 = (X4_2 + X4_4 + X4_5) / 3 = (8 + 8,8 + 7,7) / 3 = 8,16666667$$

$$C_3 = (X5_2 + X5_4 + X5_5) / 3 = (8 + 7 + 8) / 3 = 7,66666667$$

Untuk lebih jelas dapat dilihat pada tabel berikut :

Tabel 4.8 Hasil Akhir Posisi Cluster :

| Cluster | X1    | X2    | X3  | X4   | X5   |
|---------|-------|-------|-----|------|------|
| C1      | 9     | 8,3   | 9   | 9    | 8    |
| C2      | 8,175 | 8,175 | 8   | 7,35 | 6    |
| C3      | 6,83  | 7,6   | 7,6 | 8,17 | 7,67 |

Sumber : Data Olahan Penulis

6. Setelah mendapatkan centroid baru maka tahap selanjutnya adalah dilakukan perhitungan sama dengan pada tahap 3 dengan centroid baru yang didapat pada tahap 5 kemudian perulangan berhenti apabila tidak ada lagi data yang berpindah.



Keterangan :

Hasil akhir posisi cluster ditentukan berdasarkan jarak terdekat pada perhitungan akhir *cluster* dimana C1 mewakili IPA, C2 mewakili jurusan IPS dan C3 mewakili jurusan Bahasa dengan akumulasi IPA = 1 siswa, IPS = 2 siswa dan Bahasa = 3 siswa dengan 6 siswa yang dijadikan sampel dalam proses perhitungan manual yang akan di uji dengan perhitungan dengan menggunakan program.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

.Dari hasil pengujian menggunakan algoritma K-Means ini dalam penentuan jurusan siswa SMA Negeri 05 Bombana yang di ambil dari data nilai untuk 5 mata pelajaran yaitu IPA, IPS, Bahasa Inggris, Bahasa Indonesia dan Matematika dapat disimpulkan bawah :

1. Metode K-Means dapat diterapkan di dalam sistem penjurusan ini.
2. Sistem penjurusan ini dapat berfungsi dengan baik dan dapat membantu sekolah untuk menentukan siswa mana yang dapat menempati jurusan sesuai dengan kriteria yang telah ditentukan.
3. Sistem penjurusan ini dapat menampilkan data siswa kelas, nilai siswa, menentukan kriteria, menghitung jarak data dengan cendroid menggunakan metode K-Means dan mencetak hasil penjurusan sehingga mempermudah *user* untuk melihat dan mengumumkan hasil penjurusan kepada siswa.
4. Menghemat waktu yang dibutuhkan untuk memecahkan masalah serta menghasilkan solusi dengan lebih cepat dan hasilnya dapat diandalkan.
5. Mengurangi dan atau menghilangkan adanya kemungkinan kecurangan yang dilakukan oleh oknum-oknum yang tidak bertanggung jawab.

### 5.2 Saran

Adapun saran yang dapat diberikan penulis adalah bahwa bagi peneliti yang akan melakukan penelitian yang relevan dengan penelitian ini yaitu sebagai berikut :

1. Pengembangan aplikasi untuk diberikan tampilan yang lebih menarik dan *user friendly*.
2. Pengembangan aplikasi diharapkan dapat membuat laporan elektronik sehingga lebih memudahkan

siswa untuk melihat nilai dan hasil penjurusan *online*.

3. Pada sistem komputerisasi yang digunakan pada sistem penjurusan siswa ini perlu digunakan program anti virus yang residen. Hal ini dimaksudkan untuk mencegah terjadinya kerusakan pada program ataupun data-data yang lain.
4. Walaupun program aplikasi ini relatif mudah dioperasikan tapi perlu dilakukan cara-cara pengoperasian program ini, terutama bagi penanggung jawab ruangan agar sistem yang baru bisa berjalan dengan baik.

## DAFTAR PUSTAKA

- [1] Ariyanto. (2005), *Pengembangan Web di Linux dengan Apache, MySQL, dan PHP (LAMP)*, eds, Salemba Infotek, Jakarta.
- [2] Berson, Alex dan J. Smith, Stephen, (2001), *Data Warehousing, Data Mining, & OLAP*, McGraw-Hill, Singapore.
- [3] Bunafit, Nugroho, (2005), *Database Relasional dengan MySQL*, Andi Offset, Jakarta.
- [4] Direktorat Pembinaan Sekolah Menengah Atas, Permendigbud, (2017), *Model Peminatan dan Lintas Minat*. Permendigbud, Jakarta.
- [5] Husda, Nur Elfi, (2012), *Pengantar Sistem Informasi*, Badouse Media, Jakarta.
- [6] Kendal, K.E, (2011). *System Analysis Design, 8 th ed. S. Yagan Ed*, New jersey.
- [7] Kusri, (2007), *Strategi Perancangan dan Pengelolaan Basisdata*, Andi, Jakarta.
- [8] Manurung, Hartono, (2014). *Perancangan Aplikasi Penentuan Jurusan di SMA Menggunakan K-Means*, Jurnal Sistem Informasi dan Teknologi Vol.11, No.1.
- [9] Nugroho, Adi, (2005). *Analisis dan Perancangan Sistem Informasi dengan Metodologi Berorientasi Objek*. Informatika. Bandung.
- [10] Prahasta, Eddy.(2005). *Sistem Informasi Geografis Konsep-konsep Dasar*. Informatika. Bandung.
- [11] Wu X dan Kumar V, (2009), *The Top Ten Algoritma in Data Mining*, Chapman and Hall, Minnesota.