PENERAPAN DATA MINING UNTUK MEMBENTUK KELOMPOK BELAJAR MENGGUNAKAN METODE CLUSTERING DI SMPN 19 BANDUNG

Andre Catur Prasetyo

Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : andrecatur86@gmail.com

ABSTRAK

SMPN 19 Bandung adalah salah satu Sekolah Menengah Pertama Negeri di bandung. Sebagai sekolah yang menjadi pilihan favorit, SMPN 19 Bandung harus menjaga kualitas pendidikannya. Siswa kelas 9 biasanya diwajibkan mengikuti pemantapan atau *try out*. Selain pemantapan, seharusnya sekolah melakukan pembentukan kelompok belajar. Permasalahannya adalah pihak sekolah biasanya membagi kelompok hanya berdasarkan urutan absensi saja. Sehingga akan mengakibatkan siswa yang unggul berada dalam satu kelompok dengan siswa yang tertinggal dalam suatu mata pelajaran dan dikhawatirkan siswa unggul tersebut akan merasa bosan karena materi yang diberikan sudah dipahaminya diulang-ulang agar siswa yang tertinggal dapat mengejar ketertinggalannya.

Dalam *data mining*, terdapat metode yang dapat digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan kemiripan datanya, yaitu metode *clustering*. Dalam *Clustering* pun terdapat beberapa metode yang dapat digunakan, salah satunya adalah *Agglomerative Hierarchical Clustering* (AHC) dengan menggunakan algoritma *single linkage*. Proses AHC dengan menggunakan algoritma *single linkage* dimulai dengan menentukan jumlah kelompok yang akan dibentuk, menganggap seluruh data sebagai *cluster*, menghitung matriks jarak, mencari dua *cluster* terdekat lalu menggabungkannya, kemudian ulangi langkah ke-3 hingga tersisa sejumlah *cluster* yang ingin dibentuk.

Berdasarkan hasil pengujian dapat disimpulkan bahwa aplikasi Pembentuk Kelompok Belajar ini sudah membantu pihak kurikulum dalam membentuk kelompok belajar yang sesuai berdasarkan kemiripan nilai siswanya pada masing-masing kelompok.

Kata kunci: *data mining, clustering, agglomerative hierarchical clustering, single linkage,* matriks jarak

1. PENDAHULUAN

Belajar merupakan suatu kewajiban bagi siswa sekolah. Ketika duduk di bangku Sekolah Menengah Pertama (SMP) kelas 9 siswa harus memantapkan kemampuannya karena akan menghadapi ujian kelulusan yaitu Ujian Nasional (UN). SMPN 19 Bandung adalah salah satu Sekolah Menengah Pertama Negeri di bandung, tepatnya di Jalan Sadang Luhur XI Bandung. Sebagai sekolah yang menjadi pilihan favorit, SMPN 19 Bandung harus menjaga kualitas pendidikannya.

Berdasarkan observasi dan wawancara yang sudah dilakukan dengan ibu Denti selaku Wakil Kepala Sekolah Kurikulum di SMPN 19 Bandung, seluruh siswa kelas 9 pada semester genap diwajibkan mengikuti kelas tambahan yang disebut Pemantapan dan ujian try out guna mempersiapkan diri untuk menghadapi Ujian Nasional. Namun tetap saja ada siswa yang tertinggal dalam mata pelajaran tertentu karena setiap siswa memiliki potensi yang berbeda satu sama lain. Untuk dapat dinyatakan lulus, nilai siswa harus bisa melewati batas KKM (Kriteria Ketuntasan Minimal) dalam semua mata pelajaran. Selain pemantapan, seharusnya dilakukan juga pembentukan kelompok belajar.

Untuk membentuk kelompok belajar biasanya sekolah hanya membagi kelompok dengan menggunakan urutan absen. Jika pengelompokan dilakukan berdasarkan absen, besar kemungkinan siswa yang memiliki nilai tinggi di mata pelajaran akan berada dalam satu kelompok yang sama dengan siswa yang tertinggal di mata pelajaran tersebut. Hal ini akan menyebabkan siswa yang memiliki nilai tinggi akan merasa bosan karena materi yang sudah dipahaminya diulang-ulang untuk membantu siswa yang tertinggal memahami materi tersebut.

Agar siswa dapat mengejar mata pelajaran yang tertinggal, kelompok belajar harus dibentuk sesuai dengan kemiripan nilai siswa. Sehingga akan mempermudah guru untuk mengidentifikasi dimanakah kesulitan yang dialami dari setiap kelompok belajar yang terbentuk.

Maka dari itu untuk menyelesaikan masalah diatas, data mining sebagai salah satu metode untuk menggali pengetahuan dapat digunakan untuk menganalisis data akademik siswa. Salah satu metode yang dapat digunakan adalah metode clustering. Metode clustering adalah suatu teknik

dalam data mining untuk mengelompokan data yang memiliki kemiripan karakteristik dengan data lainnya. Dengan menggunakan clustering maka kelompok belajar yang dihasilkan akan beranggotakan siswa-siswa yang memiliki kemiripan dalam nilai akademik suatu mata pelajaran.

Sedangkan tujuan yang akan dicapai dalam penelitian ini adalah membantu pihak kurikulum untuk membentuk kelompok belajar berdasarkan kemiripan nilai siswa.

2. ISI PENELITIAN

2.1. Landasan Teori

Landasan teori dalam penulisan skripsi ini akan menerangkan mengenai teori-teori yang berhubungan dengan Penerapan Data Mining Untuk membentuk Kelompok Belajar di SMPN 19 Bandung.

2.1.1. *Data Mining*

Data mining adalah proses menemukan korelasi yang belum diketahui, pola dan tren dengan memilah-milah kumpulan data yang besar, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika [6].

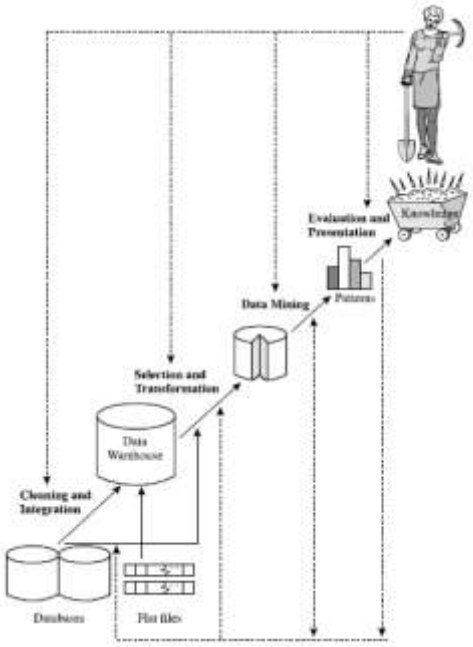
Data mining merupakan penggabungan dari pengetahuan data, dan analisa statistik yang dikembangkan dalam pengetahuan bisnis atau suatu proses menggunakan teknik statistik, matematika, kecerdasan buatan / tiruan dan machine-learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dari database yang besar [7].

Tugas-tugas data mining meliputi ekstraksi pengetahuan, arkeologi data, eksplorasi dalam pemrosesan pola data dan memanen informasi. Semua aktifitas ini dilakukan secara otomatis dan mengizinkan adanya penemuan cepat bahkan oleh non programmer [7].

2.1.2. Tahapan *Data Mining*

Data mining memiliki banyak sebutan atau istilah lain, diantaranya knowledge mining from data, knowledge extraction, data/pattern analysis, data archeology dan data dredging.

Banyak orang yang menggunakan istilah populer terhadap data mining, Knowledge Discovery from Data, atau KDD. Data mining hanya dilihat sebagai langkah penting dalam sebuah proses penemuan pengetahuan. Tahapan proses data mining digambarkan pada gambar 1.



Gambar 1 Langkah data mining dalam proses KDD [8]

Proses dalam data mining pada gambar 1 dapat dijelaskan sebagai berikut:

1. Data Selection

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (Knowledge Discovery from Data) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Preprocessing/Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

3. Transformasi

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Teknin, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretasi / Evaluasi

Pola informasi yan dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.1.3. Metode *Data Mining*

Untuk menyelesaikan penelitian data mining terdapat sebuah standar yang dapat digunakan untuk menyelesaikan penelitian data mining yaitu Cross – Industry Standard of Data Mining (CRISP-DM). CRISP-DM merupakan metodologi data mining yang komprehensif dan model proses yang memberikan siapapun (dari pemula sampai ahli data mining) blueprint secara lengkap untuk melakukan proyek data mining [12].

Berikut ini adalah tahapan-tahapan dari metodologi CRISP-DM [12]:

1. Pemahaman Bisnis

Pada awal tahap pemahaman bisnis fokus dalam pemahaman tujuan bisnis, kemudian mengubah pengetahuan tersebut ke dalam masalah data mining dan kemudian mengembangkan rencana awal yang dirancang untuk mencapai tujuan.

Tahap pemahaman bisnis dibagi menjadi beberapa langkah, antara lain:

a. Identifikasi Tujuan Bisnis

Tahap ini bertujuan untuk memahami tujuan bisnis yang ingin dicapai.

b. Sasaran Data Mining

Tujuan dari tahap ini adalah buntu menentukan kriteria sukses dari data mining.

2. Pemahaman Data

Tahap pemahaman data merupakan tahapan untuk memahami data yang berkaitan dengan penelitian yang dilakukan. Tahap pemahaman data dibagi menjadi beberapa langkah sebagai berikut:

a. Pengumpulan Data Awal

Dalam pengumpulan data awal ini digunakan untuk mengumpulkan data yang akan digunakan dalam data mining.

b. Penjelasan Data

Data yang diperoleh dari pengumpulan data awal, kemudian dijelaskan dalam tahap penjelasan data. Dalam penjelasan data akan dijelaskan format dari data, quantitas data, jumlah record dan field dalam setiap tabel, dsb.

c. Explorasi Data

Menganalisis data yang diperoleh sebelumnya antara lain analisis statistik deskriptif dan visualisasi data.

1. Analisis statistik deskriptif

Analisis statistik deskriptif ini digunakan untuk membentu terciptanya tujuan dari data mining. Analisis statistik deskriptif berkaitan dengan mencari nilai-nilai dibawah ini:

a) Nilai minimum
b) Nilai maksimum
c) Nilai rata-rata dengan menggunakan persamaan berikut ini [10]:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

d) Nilai standar deviasi dengan menggunakan persamaan berikut ini [10]:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (2)$$

2. Evaluasi Data

Dalam tahapan visualisasi data ini akan memeriksa data dengan cara menggambarkan bagaimana data yang ada, dengan cara melihat apakah terdapat missing value, ataupun outlier

a) Missing Value
b) Outlier

Outlier merupakan objek data yang menyimpang secara signifikan dari objek lainnya [10]. Metode yang dapat digunakan untuk mendeteksi outlier berdasarkan teknik statistik yaitu dengan menggunakan threshold. Untuk mendapatkan nilai threshold dapat menggunakan persamaan 3.

Batas atas = mean + (2 * standar deviasi)
Batas bawah = mean - (2 * standar deviasi) (3)

3. Persiapan Data

Tahapan persiapan data meliputi segala aktivitas pembentukan data set akhir atau data yang akan digunakan dalam pemodelan. Ada beberapa tahapan dalam tahap persiapan data ini, antara lain:

a. Pemilihan Data

Menentukan data yang akan digunakan untuk analisis berdasarkan kriteria tertentu. Pemilihan data ini meliputi pemilihan atribut ataupun pemilihan baris.

b. Pembersihan Data

Tahapan pembersihan data ini bertujuan untuk membersihkan atau menghilangkan data yang dihasilkan dalam tahapan evaluasi data sebelumnya.

c. Penyiapan Data Awal

Tahapan penyiapan data awal digunakan untuk menyiapkan data yang telah dipilih dan dilakukan pembersihan untuk digunakan dalamtahapan pemodelan.

4. Pemodelan

Dalam tahapan pemodelan ini dilakukan pemilihan model dan mengaplikasikan model yang sesuai. Langkah-langkah pemodelan adalah sebagai berikut:

a. Memilih Teknik Pemodelan

Tahapan ini digunakan untuk memilih teknik pemodelan yang sesuai dengan permasalahan dan tujuan yang ingin dicapai.

b. Pembuatan Model

Dalam tahapan ini dijelaskan mengenai teknik pemodelan yang dipilih.

c. Analisis Pengujian Model

Dalam tahapan ini model yang dipilih diaplikasikan terhadap kasus uji.

5. Evaluasi

Sebelum tahap akhir yaitu tahap pembangunan, model yang telah dibuat akan dievaluasi dan ditinjau terlebih dahulu untuk memastikan apakah model tersebut dapat mencapai tujuan atau tidak. Terdapat beberapa tahapan dalam tahap evaluasi, diantaranya adalah:

a. Mengevaluasi hasil

Tahap ini mengevaluasi model manakah yang hasilnya lebih mendekati tujuan bisnis.

b. Meninjau Proses

Melakukan peninjauan menyeluruh dari keterlibatan data mining untuk memastikan adanya faktor penting atau tugas yang terabaikan.

c. Menentukan langkah selanjutnya

6. Pembangunan

Tahap pembangunan ini merupakan tahapan implementasi untuk pembangunan aplikasi berupa representasi pengetahuan yang telah diperoleh sehingga dapat digunakan oleh pengguna.

2.1.4. Data Preprocessing

Dalam data mining kualitas dari data yang akan digunakan perlu diperhatikan. Ada beberapa faktor yang mempengaruhi kualitas data, antara lain keakuratan, keutuhan, konsistensi, aktualitas, dan penafsiran. Data preprocessing dapat memperbaiki kualitas data, sehingga dapat meningkatkan keakuratan dan efisiensi hasil dari data mining [13]. Beberapa kegiatan data preprocessing yang dilakukan dalam penelitian ini adalah sebagai berikut.

a) Menangani nilai kosong

Keberadaan nilai kosong dapat menyebabkan error pada data. Nilai kosong biasanya disebabkan oleh kesalahan manusia (human error) ketika memasukkan data. Oleh karena itu data kosong perlu ditangani dengan metode yang sesuai. Dalam penelitian ini setiap objek yang mengandung data kosong akan dihapus.

b) Menangani data noise

Data noise merupakan kesalahan acak atau variansi dalam variabel terukur. Outlier dapat direpresentasikan sebagai noise. Seperti halnya data kosong, noise juga dapat mempengaruhi kualitas data, oleh karena itu data noise perlu dihapus atau dihilangkan agar menghasilkan model yang berkualitas. Salah satu cara yang dapat dilakukan untuk membersihkan data noise adalah dengan melakukan smoothing. Salah satu teknik smoothing yang dapat dilakukan adalah metode binning.

Metode binning membagi kumpulan data ke dalam beberapa partisi atau bin. Dimulai dengan mengurutkan setiap nilai pada sebuah atribut. Kemudian data yang sudah diurutkan dibagi ke dalam beberapa partisi atau bin yang memiliki frekuensi yang sama (equal-frequency partitioning). Ada dua cara smoothing dalam metode binning, yaitu smoothing by bin means dan smoothing by bin boundaries.

Dalam smoothing by bin means dilakukan dengan mengubah setiap nilai dalam bin dengan mean dari bin tersebut. Sedangkan dalam smoothing by bin boundaries setiap nilai dalam bin diubah menjadi batas bawah (minimum) dan batas bawah (maksimum) pada setiap bin.

2.1.5. *Clustering*

Proses pengelompokan dari kumpulan objek fisik atau abstrak ke dalam kelas-kelas yang memiliki kemiripan disebut clustering. Menurut Jiawei Han dan Micheline Kamber, Sebuah cluster adalah kumpulan objek data yang memiliki kemiripan dengan objek data lainnya yang berada dalam cluster yang sama dan tidak sama dengan objek yang berada di cluster berbeda. Sebuah cluster dari objek-objek dapat dilihat sebagai satu grup dan banyak yang menganggapnya sebagai bentuk dari kompresi data [8].

Clustering juga disebut data segmentation di beberapa aplikasi karena clustering membagi-bagi set data yang besar ke dalam grup menurut kemiripannya. Dalam machine learning, clustering adalah salah satu contoh dari unsupervised learning. Tidak seperti klasifikasi, clustering tidak mengandalkan kelas yang sudah ditetapkan dan kelas training. Oleh karena itu, clustering merupakan bentuk dari learning by observation.

Menurut Santosa, pada dasarnya clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan yang lain [9].

Jiwei Han dan Micheline Kamber mengatakan bahwa secara umum, metode utama dalam clustering dikelompokan ke dalam beberapa katergori sebagai berikut [8].

1. Metode Partisi / Partition Methods
2. Metode Hirarkis / Hierarchical Methods
3. Metode Berbasis Densitas / Density-based Methods
4. Metode Berbasis Grid / Grid-based Methods
5. Metode Berbasis Model / Model-based Methods
6. Clustering high-dimensional data
7. Pengelompokan berbasis pembatas / Constraint-based clustering

Hal yang penting dalam clustering adalah mengukur kemiripan atau ketidakmiripan data yang akan dikelompokkan. Selanjutnya akan terbentuk matriks jarak yang menjelaskan kemiripan antara satu data dengan data lainnya atau jarak antara satu

data dengan data lainnya. Persamaan yang biasa digunakan untuk menghitung jarak antara dua data adalah Euclidean Distance. Berikut adalah persamaan Euclidean Distance. [5]

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{in} - x_{jn})^2} \quad (4)$$

d(i,j) = jarak antara data ke i dan data ke j
xi1 = nilai atribut ke satu dari data ke i
xj1 = nilai atribut ke satu dari data ke j
n = jumlah atribut yang digunakan

2.1.6. *Agglomerative Hierarchical Clustering*

Dalam *hierarchical clustering*, cluster yang berada di level lebih atas (*intermediate level*) dari cluster yang lain dapat diperoleh dengan cara mengkombinasikan dua buah *cluster* yang berada pada level dibawahnya. Hasil keseluruhan dari algoritma hierarchical clusteing secara grafik dapat digambarkan sebagai *tree*, yang disebut dengan dendogram.

Dalam *agglomerative hierarchical clustering*, proses *hierarchical clustering* dimulai dari cluster-cluster yang memiliki poin-poin individu yang berada di level paling bawah. Pada setiap langkahnya, dilakukan penggabungan sebuah cluster dengan cluster lainnya, dimana cluster-cluster yang digabungkan saling berdekatan atau mempunyai tingkat kesamaan yang tinggi [10].

Salah satu algoritma yang digunakan dalam agglomerative *hierarchical clustering* adalah *single linkage*. *Single linkage* terjadi apabila kelompok-kelompok digabung berdasarkan jarak yang paling dekat dari satu cluster dengan *cluster* lainnya.

Langkah-langkah dari *Agglomerative Hierarchical Clustering* dengan algoritma *single linkage* adalah sebagai berikut [11]:

1. Menentukan k sebagai jumlah *cluster* yang ingin dibentuk.
2. Setiap data dianggap sebagai *cluster*. Jika n = jumlah data dan c = jumlah *cluster*, maka c = n.
3. Menghitung matriks jarak / *similarity* / *dissimilarity* antar *cluster* (misalkan menggunakan *euclidean distance* yang terdapat dalam persamaan (4)
4. Cari dua *cluster* yang mempunyai jarak antar cluster paling kecil dan gabungkan, sehingga c = c-1.
5. Jika c > k, ulangi langkah c dan d sampai c = k.

Pada *single linkage*, kedekatan diantara dua cluster ditentukan oleh jarak terdekat (terkecil) dari dua cluster yang berbeda. Dengan cara ini maka akan dimulai dari masing-masing data sebagai cluster, kemudian mencari tetangga terdekat dengan menggunakan single linkage untuk menggabungkan dua cluster berikutnya hingga semuanya menjadi satu cluster.

Berikut ini merupakan persamaan single linkage [10]:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (5)$$

d_{UW}= jarak antara cluster u dengan cluster w
d_{VW} = jarak antara cluster v dengan cluster w

2.2. Analisis Masalah

Berdasarkan hasil observasi dan wawancara dengan pihak wakasek kurikulum, didapatkan hasil sebagai berikut:

Analisis masalah yang terdapat di SMPN 19 Bandung yaitu bahwa untuk membentuk kelompok belajar, didapatkan hasil sebagai berikut:

1. Dalam membentuk kelompok belajar, sekolah biasanya hanya menggunakan urutan absen saja.
2. Jika pengelompokan dilakukan berdasarkan absen, pembagian anggota kelompoknya tidak akan merata. Siswa yang memiliki nilai tinggi di mata pelajaran tertentu bisa saja berada dalam satu kelompok yang sama dengan siswa yang kesulitan dalam mata pelajaran tersebut.
3. Dengan pembagian kelompok yang tidak merata, akan mengakibatkan siswa yang mempunyai nilai tinggi akan merasa bosan karena materi yang diberikan diulang-ulang agar siswa yang tertinggal dapat mengejar ketertinggalan.

2.3. Pemahaman Bisnis

Tahap pemahaman bisnis merupakan tahap awal pada kerangka kerja CRISP-DM. tahap ini berfokus untuk memahami tujuan dan kebutuhan dari sudut pandang bisnis SMPN 19 Bandung. Terdapat beberapa tugas pada tahap ini, antara lain:

2.3.1. Identifikasi Tujuan Bisnis

SMPN 19 Bandung memiliki tujuan bisnis yaitu mampu membimbing siswanya agar lulus pada setiap mata pelajaran dengan melewati nilai batas KKM. Faktor penting agar tujuan bisnis SMPN 19 Bandung tercapai yaitu dengan mangadakan program-program yang dapat meningkatkan potensi akademik siswa, salah satu contohnya adalah pembentukan kelompok belajar sesuai dengan kemiripan nilai siswa.

2.3.2. Penentuan Sasaran Mining

1. Tujuan data mining

Tujuan penerapan data mining pada data siswa dalam penelitian ini yaitu untuk membentuk kelompok belajar berdasarkan nilai UTS siswa di semester ganjil yang memiliki karakteristik yang sama.

2. Kriterian sukses data mining

Kriterian sukses terhadap penelitian ini yaitu apabila mampu mengelompokan data yang memiliki karakteristik sama berdasarkan nilai UTS siswa. Ukuran kualitas kelompok yang akan diukur dengan membandingkan proses pengelompokan yang

sedang berjalan dengan proses pengelompokan dari teknik clustering pada data mining.

2.4. Pemahaman Data

Tahapan pemahaman merupakan tahapan kedua dari kerangka kerja CRISP-DM. terdapat beberapa tugas dalam tahap ini yaitu:

2.4.1. Pengumpulan Data Awal

Data yang digunakan dalam penelitian yaitu data nilai ujian tengah semester (UTS) ganjil siswa kelas 9E, 9F, dan 9G Tahun 2015. Data nilai UTS digunakan karena murni hasil ujian siswa tanpa remedial.

Data nilai yang akan digunakan berupa file berformat *.xls atau .xlsx. Selain itu data yang digunakan pada penelitian ini hanya data yang berada dalam tanggung jawab SMPN 19 Bandung.

2.4.2. Penjelasan Data

Data nilai yang digunakan memiliki 7 atribut, 3 atribut pertama menjelaskan profil siswa, sedangkan 4 atribut lainnya menjelaskan nilai Ujian Tengah Semester genap siswa. Jumlah data yang terdapat pada data nilai sebanyak 107 record.

atribut yang terdapat pada data nilai UTS siswa kelas 9E, 9F, dan 9G tahun dijelaskan pada tabel 1 di bawah ini.

Tabel 1 Penjelasasn Atribut Data

No	Atribut	Tipe Data	Keterangan
1.	No Induk	String	Merupakan kode identitas siswa yang bersifat unik
2.	Nama	String	Merupakan nama siswa
3.	Jenis Kelamin	String	Jenis kelamin siswa
4.	Nilai UTS B.Ind	Numeric	Nilai UTS mata pelajaran B. Indonesia semester genap siswa
5.	Nilai UTS B.Ing	Numeric	Nilai UTS mata pelajaran B. Inggris semester genap siswa
6	Nilai UTS Mat	Numeric	Nilai UTS mata pelajaran Matematika semester genap siswa
7.	Nilai UTS IPA	Numeric	Nilai UTS mata pelajaran IPA semester genap siswa

2.4.3. Eksplorasi Data

Tahapan eksplorasi data dapat membantu tujuan dari data Mining, dalam penelitian ini eksplorasi data meliputi analisis deskriptif dan evaluasi kualitas data. Data yang akan digunakan untuk tahap eksplorasi data ini adalah data nilai UTS siswa kelas 9E. Berikut ini hasil eksplorasi data terhadap data nilai siswa yaitu:

1. Analisis deskriptif

Analisis deskriptif ini dilakukan pada atribut Nilai UTS B.Ind, Nilai UTS B.Ing, Nilai UTS Mat, dan Nilai UTS IPA. Analisis dilakukan terhadap atribut tersebut karena atribut pengelompokan yang akan dilakukan berdasarkan nilai UTS siswa.

Dalam tahapan ini akan dilakukan tahapan analisis terhadap data nilai UTS kelas 9E tahun 2015, analisis tersebut dilakukan pada setiap atribut nilai UTS. Berikut ini analisis yang akan dilakukan antara lain:

a. Nilai minimum
   Berikut nilai minimum UTS dari setiap mata pelajaran:
   Min Nilai UTS B.Ind = 50
   Min Nilai UTS B.Ing = 68
   Min Nilai UTS Mat = 20
   Min Nilai UTS IPA = 70
b. Nilai maksimum
   Berikut nilai maksimum UTS dari setiap mata pelajaran:
   Max Nilai UTS B.Ind = 80
   Max Nilai UTS B.Ing = 88
   Max Nilai UTS Mat = 100
   Max Nilai UTS IPA = 100
c. Nilai rata – rata
   Cara untuk mendapatkan nilai rata – rata dapat menggunakan persamaan (1). Maka dengan menggunakan persamaan (1) didapatkan hasil sebagai berikut:
   1) Rata-rata nilai UTS B.Ind

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{2470}{36} = 68.61$$

   2) Rata-rata nilai UTS B.Ing

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{2850}{36} = 79.17$$

   3) Rata-rata nilai UTS Mat

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{2575}{36} = 71.53$$

   4) Rata-rata nilai UTS IPA

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{3101}{36} = 86.14$$

d. Standar Deviasi
Dengan menggunakan persamaan (2) didapatkan hasil sebagai berikut:
Tabel 2 Hasil Analisis Deskriptif

No	Mata Pelajaran	Nilai min	Nilai Max	Nilai rata – rata	Standar Deviasi
1	B. Indonesia	50	80	68.61	8.05
2	B. Inggris	68	88	79.17	4.90
3	Matematika	20	100	71.53	24.57
4	IPA	73	100	86.14	7.75

2. Evaluasi Kualitas Data
Evaluasi kualitas data dilakukan terhadap atribut Nilai UTS B.Ind, Nilai UTS B.Ing, Nilai UTS Mat, dan Nilai UTS IPA. Berikut ini merupakan hasil evaluasi dari atribut tersebut:

a. Terdapat missing value pada data kelas 9E yaitu pada atribut no_induk. Missing value pada atribut no induk akan diganti menjadi menjadi PINDAHAN1, PINDAHAN2, dan seterusnya.
b. Dari data nilai tersebut akan dilihat apakah outlier atau tidak, dan berikut ini merupakan cara yang dapat dilakukan untuk melihat data outlier dengan menggunakan persamaan (3). Maka diperoleh batas atas dan batas bawah seperti yang dapat dilihat pada tabel 3.

Tabel 3 Batas Atas dan Batas Bawah *Outlier*

No	Mata Pelajaran	Batas Atas	Batas Bawah
1	B. Indonesia	84.71	52.51
2	B. Inggris	88.97	69.37
3	Matematika	120.67	22.39
4	IPA	101.94	70.34

2.5. Persiapan Data

Tahapan ini mencakup semua kegiatan yang diperlukan untuk membangun dataset akhir (data yang akan dimasukkan ke dalam modeling tools) dari data mentah awal, terdapat beberapa tugas dalam tahap ini, yaitu:

2.5.1. Pemilihan Data

Pemilihan data memiliki tugas meliputi pemilihan atribut dan baris. Atribut yang dipilih untuk melakukan proses pengelompokan (clustering) antara lain, No Induk, Nilai UTS B.Ind, Nilai UTS B.Ing, Nilai UTS Mat, Nilai UTS IPA.

2.5.2. Pembersihan Data

Pada proses ini dilakukan proses pembersihan data, berdasarkan hasil verifikasi kualitas diperoleh hasil sebagai berikut:

1. Missing value ditemukan pada atribut no_induk. Missing value pada atribut no induk akan diganti menjadi menjadi PINDAHAN1, PINDAHAN2, dan seterusnya.
2. Nilai yang bersifat outlier akan tetap diproses, untuk melakukan pengelompokan siswa akan digunakan algoritma yang tahan terhadap adanya outlier, yaitu dengan melakukan proses smoothing terhadap data menggunakan metode binning.

2.5.3. Penyiapan Data Awal

Pada proses penyiapan data awal akan disiapkan data yang akan digunakan dalam tahapan pemodelan. Data yang akan digunakan dalam tahapan pemodelan adalah data nilai kelas 9E tahun 2015 yang sudah di-*smoothing*.

2.6. Pemodelan

Tahapan selanjutnya dari kerangka kerja CRISP-DM adalah pemodelan, tahap ini memiliki tugas untuk pemilihan teknik pemodelan, pembuatan model, serta pengujian model.

2.6.1. Teknik Pemodelan

Teknik pemodelan yang digunakan sesuai dengan tujuan awal yaitu untuk membentuk kelompok belajar. Model yang akan diuji yaitu algoritma Agglomerative Hierarchical Clustering (AHC).

2.6.2. Pembuatan Model

Tahap ini menjelaskan mengenai mekanisme dari model yang akan digunakan serta asumsi awal dari model. Model algoritma Agglomerative hierarchical clustering memiliki tahapan sebagai berikut:

1. Menentukan k sebagai jumlah cluster yang akan dibentuk.
2. Setiap data dianggap sebagai cluster. Jika n = jumlah data dan c = jumlah cluster, maka c = n.
3. Menghitung matriks jarak / similarity / dissimilarity antar cluster (misalkan menggunakan euclidean distance yang terdapat dalam persamaan 1)
4. Cari dua cluster yang mempunyai jarak antar cluster paling kecil dan gabungkan, sehingga c = c-1. Jika cluster (n, n) dan euclidean distance = 0, maka lanjutkan pencarian.
5. Jika c > k, ulangi langkah ke-3 dan ke-4 sampai c = k.

2.6.3. Analisis Pengujian Model

Data yang akan digunakan untuk melakukan pengelompokan ini adalah data nilai UTS siswa kelas 9E tahun 2015. Jumlah data yang digunakan adalah sebanyak 36 record.

Kasus yang akan diuji dengan menggunakan algoritma AHC ini adalah sebagai berikut:

1. Menentukan jumlah cluster yang akan dibentuk

Jumlah kelompok belajar yang akan dibentuk adalah sebanyak tiga kelompok berdasarkan jumlah kelompok yang biasa dibentuk yaitu tiga. Jika k = jumlah cluster yang akan dibentuk, maka k = 3.

2. Ubah setiap data menjadi cluster

Inisialisasi cluster awal adalah sebanyak 36 cluster karena data(n) = cluster(c), kemudian dihitung jarak antara cluster dengan cluster yang ada dengan menggunakan persamaan Euclidean Distance (persamaan 4).

3. Menghitung jarak antar data

Dalam tahap ini masing – masing cluster akan dihitung jarak dengan dirinya sendiri dan jarak dengan cluster lainnya. Berikut ini merupakan proses perhitungan jarak antar cluster dengan menggunakan persamaan Euclidean Distance (persamaan 4).

$$d(1,1) = \sqrt{(72-72)^2 + (79-79)^2 + (95-95)^2 + (93-93)^2} = 0$$
$$d(1,2) = \sqrt{(72-55)^2 + (79-76)^2 + (95-30)^2 + (93-80)^2} = 68.50$$
$$d(1,3) = \sqrt{(72-67)^2 + (79-79)^2 + (95-95)^2 + (93-90)^2} = 5.39$$
$$\vdots$$
$$d(36,36) = \sqrt{(72-72)^2 + (71-71)^2 + (81-81)^2 + (73-73)^2} = 0$$

Dengan cara yang sama yaitu dengan menggunakan persamaan Euclidean Distance dilakukan perhitungan terhadap seluruh data. Dari proses perhitungan Euclidean Distance maka dihasilkan matriks jarak.

4. Mencari dua cluster yang paling berdekatan

Dalam tahapan ini akan dihitung jarak antara dua cluster dengan menggunakan metode single linkage (persamaan 5), dengan metode ini akan mencari jarak yang paling dekat dari dua cluster.

Iterasi ke-1:

$$d_{(1,2)3} = \min\{d_{(1,3)}, d_{(2,3)}\} = \min\{68.50, 5.39\} = 5.39$$

sampai dengan

$$d_{(u,v)n} = \min\{d_{(u,n)}, d_{(v,n)}\}$$

Dengan metode single linkage jarak yang paling dekat, dari setiap iterasi yang dilakukan akan menggabungkan dua cluster, sehingga jumlah iterasi yang akan dihasilkan dari jumlah 36 data dan cluster yang akan dibentuk sebanyak k=3 cluster maka akan menghasilkan sebanyak 33 iterasi. Jika cluster (n, n) dan euclidean distance = 0, maka nilai euclidean distance tersebut tidak dinyatakan sebagai yang terkecil dan lanjutkan pencarian. Misalkan cluster (1, 1), cluster (2,2), cluster (3,3), dan seterusnya yang memiliki nilai euclidean distance = 0. Dari iterasi yang pertama jarak yang paling dekat adalah cluster 13 dengan 23 dengan nilai jarak yaitu 2.00, maka kedua cluster tersebut digabungkan menjadi satu cluster dan jumlah cluster baru yang terbentuk sebanyak 35 cluster.

5. Jika c > k, ulangi langkah ke-3 dan ke-4

Langkah ke-3 dan ke-4 terus dilakukan dan akan berhenti ketika jumlah cluster yang terbentuk (c) = jumlah cluster yang ingin dibentuk (k) atau c = k = 3. Untuk mendapatkan tiga cluster, dalam kasus ini perlu dilakukan sebanyak 33 iterasi.

Dari proses pengelompokan dengan menggunakan algoritma AHC didapatkan hasil sebagai berikut:

a. Dengan menggunakan jumlah cluster sebanyak tiga cluster dari jumlah data yang digunakan sebanyak 36 data, 7 data berada di cluster 1, 22 data berada di cluster 2 dan 7 data di cluster 3.
b. Dengan menggunakan metode single linkage pengelompokan akan cenderung berkumpul di satu cluster karena setiap iterasi yang dilakukan mencari nilai terkecil dari setiap cluster, dan nilai tersebut digunakan untuk dibandingkan dengan cluster lain.
c. Cluster 1 dan cluster 3 memiliki nilai rata-rata yang cukup rendah dalam mata pelajaran bahasa Indonesia dan matematika.
d. Cluster 2 memiliki nilai rata-rata cukup rendah pada mata pelajaran Bahasa Indonesia.

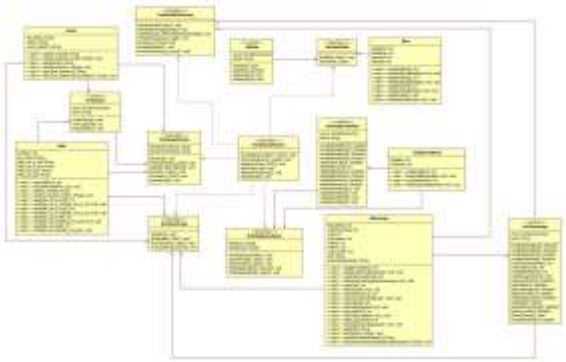
2.6. Diagram *Use Case*

Use case atau diagram *use case* merupakan pemodelan untuk kelakuan (*behavior*) sistem yang akan dibuat. Diagram *use case* yang terdapat pada sistem yang akan dibangun terdiri dari satu aktor dan sepuluh *use case*.



Gambar 2 Diagram *Use Case* Sistem Pembentuk Kelompok Belajar SMPN 19 Bandung

2.7. *Class Diagram*



Gambar 3 Class Diagram

2.8. Skema Relasi

Skema relasi merupakan rangkaian hubungan antara dua tabel atau lebih pada sistem basis data. Berikut ini adalah skema relasi pada sistem pembentukan kelompok belajar di SMPN 19 Bandung:



Gambar 4 Skema Relasi

2.9. Perancangan Struktur Menu

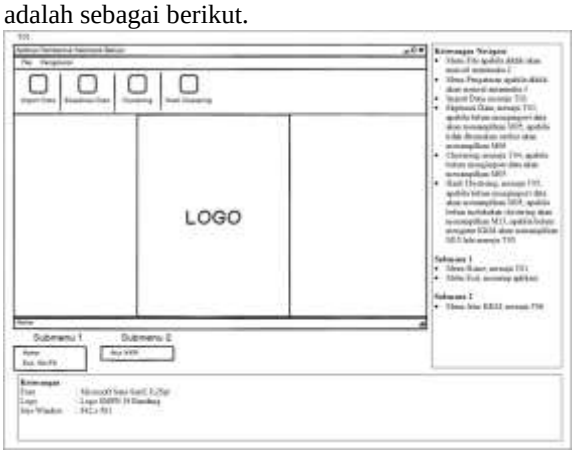
Perancangan struktur menu pada aplikasi yang akan dibangun hanya untuk satu user / pengguna. Berikut adalah perancangan struktur menu pada aplikasi yang akan dibangun.



Gambar 5 Struktur Menu

2.10. Perancangan Antarmuka

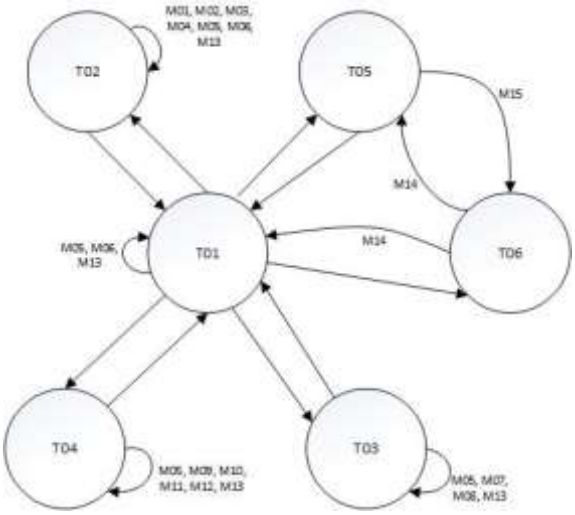
Perancangan Antarmuka adalah rancangan tampilan sistem yang akan dibangun yang berfungsi sebagai media komunikasi yang digunakan sebagai sarana untuk berinteraksi antara sistem dengan pengguna. Sistem yang dibangun diharapkan menyediakan tampilan yang mudah dipahami dan mudah digunakan oleh pengguna. Perancangan antarmuka pada sistem yang dibangun adalah sebagai berikut.



Gambar 6 AntarmukaMenu Utama

2.11. Jaringan Semantik

Jaringan Semantik adalah keterhubungan dari navigasi suatu menu dari satu halaman ke halaman yang lainya. Jaringan semantik dari sistem yang akan dibangun dapat dilihat pada gambar 7.



Gambar 7 Jaringan Semantik

2.12. Pengujian Sistem

Pengujian sistem dilakukan terhadap fungsionalitas sistem untuk menilai apakah aplikasi yang dibangun sudah memenuhi kriteria atau terdapat ketidaksesuaian pada fungsinalitas sebagaimana yang terdapat pada perancangan sistem. Pengujian yang digunakan untuk menguji perangkat lunak yang akan dibangun adalah metode pengujian black box.

Rencana pengujian yang akan dilakukan adalah dengan cara menguji aplikasi yang dibangun secara black box dan beta. Pengujian black box terfokus pada pengujian funsionalitas perangkat lunak yang akan dibangun.

1. Pengujian *Black Box*

Pengujian black box difokuskan pada persyaratan fungsional perangkat lunak yang dibangun.

Pengujian dilakukan dengan menguji setiap proses untuk kemungkinan kesalahan yang terjadi.

Berdasarkan hasil pengujian dengan kasus sampel uji yang telah dilakukan dapat ditarik kesimpulan bahwa seluruh fungsi berjalan dengan baik dan sesuai dengan yang diharapkan.

2. Pengujian Beta

Pengujian beta merupakan pengujian yang dilakukan langsung di lingkungan sebenarnya. Pengguna memberikan penilaian terhadap perangkat lunak yang dibangun melalui metode wawancara. Berdasarkan hasil wawancara tersebut maka dapat ditarik kesimpulan apakah perangkat lunak yang dibangun sesuai dengan tujuan atau tidak.

Wawancara dilakukan untuk perangkat lunak yang dibangun, yaitu staf bagian kurikulum. Penelitian ini dilakukan di SMPN 19 Bandung agar dapat mengetahui sejauh mana perangkat lunak yang dibangun dapat menjadi solusi dan penyelesaian masalah yang telah dijelaskan sebelumnya.

Berdasarkan pengujian beta yang telah dilakukan sebelumnya, dapat disimpulkan bahwa aplikasi yang dibangun dapat membantu pihak bagian kurikulum untuk membentuk kelompok belajar yang sesuai berdasarkan kemiripan nilai siswa.

3. PENUTUP

Pada bagian ini, menjelaskan tentang kesimpulan yang berisi hasil-hasil yang diperoleh setelah dilakukan analisis, desain, dan implementasi dari perancangan perangkat lunak yang dibangun dan telah dikembangkan serta saran-saran yang akan memberikan catatan penting dan kemungkinan perbaikan yang perlu dilakukan untuk pengembangan perangkat lunak sebelumnya.

3.1. Kesimpulan

Hasil dari penelitian yang dilakukan diperoleh kesimpulan bahwa aplikasi yang dibangun membantu pihak bagian kurikulum untuk membentuk kelompok belajar yang sesuai dengan kemiripan nilai siswa.

3.2. Saran

Ada beberapa saran yang dapat dilakukan untuk mengembangkan aplikasi pembentuk kelompok belajar ini, antara lain:

1. Penambahan fitur pengaturan tahun ajaran agar meningkatkan keakuratan clustering dan memudahkan user untuk menyaring data hasil clustering yang akan digunakan apabila data yang dikelompokan campuran dari tahun ajaran yang berbeda.
2. Pengembangan pengelompokan dapat dikembangkan lagi dengan menggunakan algoritma lain sehingga user dapat membandingkan hasil pengelompokannya dengan algoritma Single Linkage.

DAFTAR PUSTAKA

[1] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, CRISP-DM 1.0, SPSS Inc, 2000.

[2] Kadir and Abdul, Pengenalan Sistem Informasi, Yogyakarta: ANDI, 2003.

[3] G. B. Davis, Management Information System: Conceptual Foundations, Structure, and Development, Auckllland: McGraw-Hill International Book Company, 1974, p. 32.

[4] J. G. Burch and F. R. Strater, Information System: Theory and Practice, California: Hamilton Publishing Company, 1974, p. 23.

[5] G. R. Terry, Office Management and Control, 4th ed., Ilinois: Richard D. Irwin Inc., 1962, p. 21.

[6] D. T. Larose, Discovering Knowledge in Data, Hoboken: WILEY-INTERSCIENCE, 2005.

[7] S. Kurniawan and T. Hidayat, "PENERAPAN DATA MINING DENGAN METODE INTERPOLASI UNTUK MEMPREDIKSI MINAT KONSUMEN ASURANSI (Studi Kasus Asuransi Metlife)," vol. 5, pp. 113-128, 2007.

[8] J. Han and M. Kamber, Data Mining: Concepts and Technique, 2nd ed., San Fransisco: MORGAN KAUFMANN, 2006.

[9] B. Santosa, Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.

[10] G. A. Pradnyana and N. A. S. ER, "Perancangan dan Implementasi Automated Document Integration dengan Menggunalan Algoritma Complete Linkage Aggglomerative Hierarchical Clustering," Jurnal Ilmu Komputer, vol. 5, no. 2, pp. 1-10, September 2012.

[11] R. Handoyo, R. R. M and S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means Pada Pengelompokan Dokumen," JSM STMIK Mikroskil, vol. 15, no. 2, pp. 73-82, Oktober 2014.

[12] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data," Journal of Data Warehousing, vol. 5, no. 4, p. 14, 2000.

[13] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques 3rd Edition, San Fransisco: Morgan Kauffman.

[14] R. Miles and K. Hamilton, Learning UML 2.0, Sebastopol: O'Reilly Media Inc., 2006.