

BAB 2

LANDASAN TEORI

2.1. Data Mining

2.1.1. Pengertian Data Mining

Kehadiran *data mining* dilatar belakangi dengan problema data *explosion* yang dialami akhir-akhir ini dimana banyak organisasi telah mengumpulkan data sekian tahun lamanya yang ada disekolahan, dan akhirnya data tersebut hilang (data induk siswa, data transkrip nilai, data transaksi pembayaran sekolah, dsb). Hampir semua data tersebut dimasukkan dengan menggunakan tulisan tangan atau ditulis secara manual sehingga data-data tersebut bisa saja hilang atau terselip.

Data mining adalah sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis atau semi otomatis dari sekumpulan data dalam jumlah besar. *Data mining* hadir dianggap sebagai bagian dari *Knowledge Discovery in Database* (KDD) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data. KDD terdiri dari beberapa langkah [SAN07] yaitu:

1. Pembersihan data (membuang *noise* dan data yang tidak konsisten)
2. Integrasi data (penggabungan data dari beberapa sumber)
3. Seleksi data (memilih data yang relevan yang akan digunakan untuk analisa)
4. Data mining
5. Evaluasi model
6. Presentasi pengetahuan dengan teknik visualisasi

2.1.2. Metode Data Mining

Pada umumnya metode data mining dapat dikelompokkan ke dalam dua kategori, yaitu: *deskriptif* dan *prediktif*. Metode *deskriptif* bertujuan untuk mencari pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik dari

data. Metode *clustering* menggunakan ciri-ciri tertentu dari data untuk melakukan pengelompokan data [KUS09]. Metode-metode yang ada dalam *data mining* adalah sebagai berikut:

1. *Classification*

Klasifikasi merupakan proses untuk menemukan sekumpulan model yang menjelaskan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah obyek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test uji*) digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai dari suatu obyek data.

2. *Clustering*

Pengelompokan (*Clustering*) merupakan proses untuk melakukan segmentasi. Digunakan untuk melakukan pengelompokan secara alami terhadap atribut suatu set data. Termasuk kedalam *undersurvised task*. Metode inilah yang digunakan dalam tugas akhir ini.

3. *Assosiation*

Tujuan dari metode ini yaitu untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang terhubung kuat satu dengan yang lainnya. Sebagai contoh *association analysis* dapat digunakan untuk menentukan produk yang datang dibeli secara bersamaan oleh banyak pelanggan, atau bisa juga disebut dengan *market based analysis*.

4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diprediksi berupa nilai yang *kontinyu*.

5. *Forecasting*

Prediksi (*Forecasting*) berfungsi untuk melakukan prediksi kejadian yang akan datang berdasarkan data sejarah yang ada.

6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari *data diskrit*. Sebagai contoh adalah menemukan kelompok gen dengan tingkat ekspresi yang mirip.

7. *Deviation Analysis*

Tujuan dari metode ini adalah untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut sebagai *outlier detection*. Sebagai contoh adalah apakah sudah terjadi penipuan terhadap pengguna kartu kredit dengan melihat catatan transaksi yang tersimpan dalam basis data perusahaan tersebut.

2.2. Clustering

Analisis kluster atau *clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok yang lain [KUS09].

Potensi *clustering* dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lanjut dalam berbagai aplikasi secara luas seperti *klasifikasi*, pengolahan gambar, dan pengenalan pola. *Clustering* dapat diterapkan ke dalam data yang *kuantitatif* (numerik), *kualitatif* (kategorikal), atau kombinasi keduanya [KUS09]. Data dapat merupakan hasil pengamatan dari suatu proses. *Cluster* secara umum merupakan wujud himpunan bagian dari suatu himpunan data dan metode *clustering* dapat diklasifikasi berdasarkan himpunan bagian yang dihasilkan: apakah *fuzzy* atau *crisp* (*hard*).

Pada Penelitian Tugas Akhir ini akan dibahas mengenai penggunaan metode *Fuzzy K-Means* dengan pengelompokan potensi siswa.

2.3 Algoritma K-Means

K-Means merupakan algoritma *clustering* yang berulang-ulang. Algoritma K-Means dimulai dengan pemilihan secara acak N , N disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai N secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) [SAN07].

Dari beberapa teknik klastering yang paling sederhana dan umum dikenal adalah klastering *K-mean*. Dalam teknik ini kita ingin mengelompokkan obyek kedalam N kelompok atau klaster. Untuk melakukan klaustering, nilai N harus ditentukan terlebih dahulu. Biasanya user atau pemakai sudah mempunyai informasi awal tentang obyek yang sedang dipelajari, termasuk beberapa jumlah klaster yang paling tepat. Secara detail kita bisa menggunakan ukuran ketidakmiripan untuk mengelompokkan obyek kita. Ketidakmiripan bisa diterjemahkan dalam konsep jarak. Jika jarak dua obyek atau data titik cukup dekat maka dua obyek itu mirip. Semakin dekat berarti semakin tinggi kemiripannya. Semakin tinggi nilai jarak semakin tinggi ketidakmiripannya.

Dalam penelitiannya mengungkapkan langkah-langkah pengerjaan algoritma K-means [SAN07] yaitu :

1. Penentuan pusat cluster awal

Dalam penentuan nilai N buah pusat cluster awal dilakukan pembangkitan bilangan random yang mempresentasikan urutan data input.pusat awal cluster didapatkan dari data sendiri bukan

denagn menentukan titik baru, yaitu dengan merandom pusat awal dari data

2. Perhitungan jarak dengan pusat cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan rumus *Euclidian distance*. Algoritma perhitungan jarak data dengan pusat cluster

- a. Ambil nilai data dan nilai pusat cluster.
- b. Hitung *Euclidian distance* data dengan tiap pusat cluster.

3. Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini akan menunjukkan data tersebut berada dalam suatu kelompok dengan pusat cluster terdekat. Algoritma pengelompokan data :

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

4. Penentuan pusat cluster baru

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

Algoritma penentuan pusat cluster baru yaitu:

- a. Cari jumlah anggota tiap cluster.

- b. Hitung pusat baru dengan rumus.

$$\text{Pusat cluster baru} = \frac{x_1+x_2+x_3,\dots,x_n}{\text{jumlah}+1}$$

Dimana :

$x_1, x_2, x_3, \dots, x_n$ adalah anggota cluster.

Berikut ini adalah uraian dari perancangan algoritma K-means untuk menentukan pengelompokan potensi atau nilai siswa.

2.4 Profil Instansi

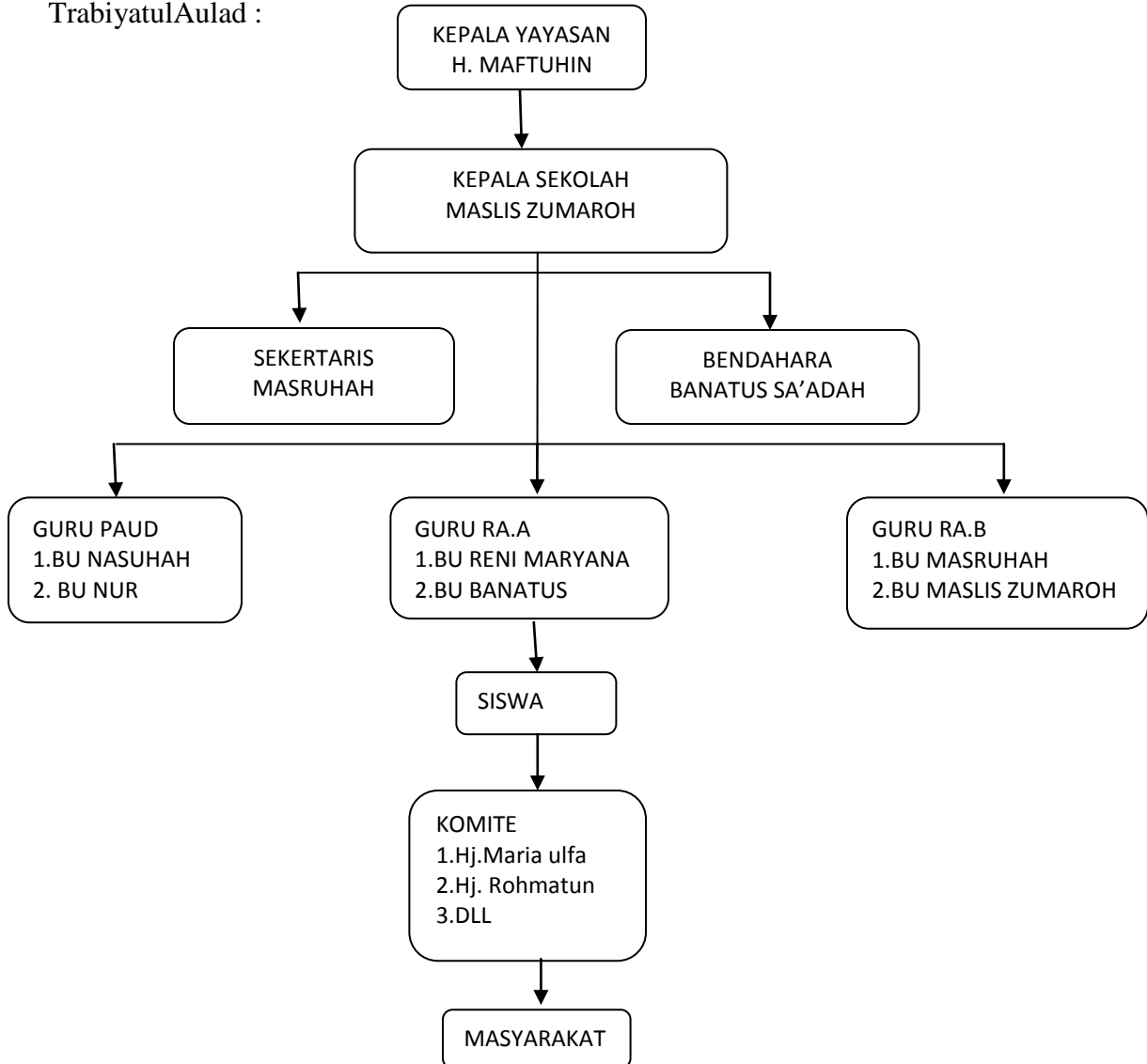
2.4.1 Sejarah sekolah

Pada awal berdirinya ditahun 2008 sekolah ini hanya bersifat temporer dengan status terbuka dan belum terakreditasi, beroperasi di atas bangunan yang hanya seluas 5,6 X 6 M2 dan berjumlah murid tidak lebih dari 5 orang. Seiring berjalannya waktu, jumlah murid dari sekolah ini pun makin bertambah, dan akhirnya pengelola pun merasa membutuhkan gedung yang lebih luas dan layak untuk tempat belajar para muridnya. Dengan bantuan dari warga desa dan bantuan dari pemerintah baik dari Departemen Agama(DEPAK) maupun dari Departemen Pendidikan (DIKNAS), maka dibangunlah sebuah gedung tambahan yang dibangun di dekat area sekolah tersebut seluas 6 X 12M2.

Karena prestasi dan kualitasnya yang sudah cukup baik, maka pada tahun 2010 sekolah ini diresmikan oleh Departemen Agama dengan nama RA. Tarbiyatul Aulad. Dengan jumlah murid pada saat itu adalah 59 orang yang terdiri dari 2 kelas yaitu kelas RA A berjumlah 29 dan yang RA B 26. Dan selalu mendapatkan piala berturut-turut selama 3 tahun kebelakang dengan kategori siswa berprestasi dalam lomba HAN (Hari Anak Nasional) yang diselenggarakan di kecamatan Cerme.

Berikut adalah struktur organisasi sekaligus pembina sekolah RA.

TrabiyatulAulad :



(Gambar 2.1 Struktur Organisasi RA Tarbitayul Aulad)

Sekolah ini beralamatkan di Desa Wedani Kecamatan Kecamatan Cerme Kabupaten Gresik, dengan status gedung “Milik Sendiri”, nomor statistik 101231720036 dan izin pendirian WJ/6B/PP004/3987/92. Di sekolah ini juga menyediakan beberapa ekstrakurikuler bagi para muridnya, yaitu menari, calistung, bahasa inggris, mewarnai dengan perpaduan warna yang baik, sempoa, dan computer yang masing-masing dari kegiatan ekskul tersebut telah meraih prestasi yang cukup baik hingga saat ini.

2.4.2 Deskripsi Pekerjaan Instansi

Deskripsi pekerjaan (*Job description*) adalah uraian dari setiap tugas atau tanggung jawab yang diemban oleh setiap anggota dari suatu organisasi, agar fungsi dari masing-masing anggota menjadi jelas ruang lingkupnya.

Berikut adalah *jobdesc* yang ada di RA. Tarbiyatul Aulad :

1. Kepala Sekolah :

- a. Monitoring kegiatan belajar mengajar (KBM)
- b. Pengambilan keputusan
- c. Menghadiri segala kegiatan yang berhubungan dengan prosedur KBM-nya.

2. Guru :

- a. Mengajar
- b. Membuat laporan perencanaan pembelajaran
- c. Monitoring siswa didiknya
- d. Membuat laporan akhir siswa

3. Tata Usaha (TU)

- a. Pembukaan surat masuk dan surat keluar

- b. Mengurus seluruh proses administrasi sekolah, seperti ; PSB, iuran bulanan, dan kegiatan-kegiatan lainnya yang membutuhkan biaya.

2.4.3 Pengertian Sekolah Taman Kanak-kanak

Taman Kanak-Kanak (TK) termasuk dalam jenjang Pendidikan Anak Usia Dini (PAUD), yakni usia 6 tahun atau di bawahnya yang dikemas dalam bentuk pendidikan formal. Program pembelajaran TK lebih ditekankan kepada pemberian rangsangan pendidikan guna membantu pertumbuhan dan perkembangan jasmani dan rohani agar anak memiliki kesiapan dalam memasuki pendidikan lebih lanjut. Masa belajar seorang murid di TIK tergantung pada tingkat kecerdasannya yang dinilai dari rapor per semesternya. Untuk lulus dari tingkat program di TK selama dua (2) tahun, yakni :

1. TK Nol Kecil selama 1 tahun
2. TK Nol Besar selama 1 tahun

Usia rata-rata minimal anak untuk dapat memulai pendidikan di TK berkisar 4-5 tahun, sedangkan usia rata-rata untuk lulus dari TK berkisar 6-7 tahun. Setelah lulus dari TK, murid akan melanjutkan program pendidikan ke tingkat Sekolah Dasar atau yang biasa disingkat SD.

2.4.4 Pengertian *Raudhatul Athfal*

Jenjang pendidikan anak usia dini (yakni usia 6 tahun atau di bawahnya) dalam bentuk pendidikan formal, di bawah pengelolaan Kementerian Agama RA setara dengan taman kanak-kanak (TK), di mana kurikulumnya ditekankan pada pemberian rangsangan pendidikan untuk membantu pertumbuhan dan perkembangan jasmani dan rohani agar anak memiliki kesiapan dalam memasuki pendidikan lebih lanjut.

2.4.5 Teknik Pembelajaran Siswa Sekolah Taman Kanak-kanak

Teknik pembelajaran di TK yaitu dengan memberi kesempatan anak-anak belajar sesuai dengan usia tiap tingkatannya, antara lain :

- a. Bernyanyi
- b. Membaca
- c. Berhitung
- d. Budi bahasa
- e. Agama
- f. Dan berbagai macam keterampilan lainnya.

Tujuan dari pembelajaran tersebut diatas adalah untuk menciptakan daya cipta kanak-kanak serta memacunya untuk belajar mengenai berbagai ilmu pengetahuan yang dirancang sebagai upaya untuk menumbuhkembangkan daya pikir dan peranan anak kecil, yang dikemas dalam bentuk belajar sambil bermain.

2.4.6 Kriteria yang digunakan dalam pembelajaran

Kriteria yang digunakan dalam pembelajaran di RA Tarbiyatul Aulad yaitu sebagai berikut :

- a. Diskusi dan ahlak prilaku social emosional
Dalam pelajaran ini siswa mampu mengendalikan emosi diri dalam prilaku yang baik dan buruk serta mampu berdiskusi dengan teman yang lain. Selain itu siswa juga diajarkan saling menyayangi satu sama yang lain dan mencintai alam sekitar.
- b. Hafalan surat pendek dan do'a-do'a
Dalam pelajaran ini siswa mampu menghafalkan surat pendek dan do'a sehari-hari. Dan menerapkannya dilingkungan sekolah dan dirumah.
- c. Agama, ahlak dan bahasa arab

Pelajaran ini siswa diajari agama, ahlak yang baik dan benar serta diajari menulis bahasa arab. Selain itu siswa agar selalu mengagumi dan mensyukuri semua yang diberikan oleh Allah.

d. Bahasa

Dalam pelajaran ini siswa diajari cara berbahasa yang baik serta menulis huruf. Selain berbahasa dan menulis siswa juga diajari berbagai macam bahasa yang mudah dipahami dan mudah diingat seperti bahasa inggris, bahasa arab, dan bahasa jawa.

e. Kognitif math

Pelajaran ini siswa diajari cara berhitung dan mengenal angka-angka. Dalam mengenalkan angka siswa diajak bermain dan bernyanyi angka-angka agar siswa mudah mengingatnya.

f. Motorik kasar

Dalam pelajaran ini siswa diajak berolah raga dan melakukan kegiatan yang berhubungan dengan olah raga. Di pelajaran ini siswa juga diajarkan bertanam buah-buahan, bungah dan menanam biji-bijian serta cara merawat tanaman tersebut agar bisa tumbuh dengan subur dan indah.

g. Motorik halus

Pelajaran ini siswa mampu berkreasi dan menggambar gambar yang sudah di sediakan oleh guru dan dapat menggambar sendiri sesuai dengan imajinasinya siswa masing-masing. Tidak hanya menggambar saja siswa juga mampu mengexplor kan diri sesuai dengan keinginannya.

h. Bahasa inggris

Dalam pelajaran ini siswa diajari kosa kata bahasa inggris yang mudah dipahami dan dihafalkan.

i. Kognitif sains

Dalam pelajaran ini siswa diajak bermain sains dan berexperiment berbagai kegiatan yang sudah disediakan oleh ibu guru yang baik dan benar. Seperti contoh mencampurkan warna kemudian berubah jadi apa, membedakan benda yang bisa mengapung, tenggelam.

2.4 Penelitian sebelumnya

Penelitian sebelumnya dilakukan berorientasi obyek yaitu berdasarkan data-data yang sudah tersimpan seperti nilai-nilai bidang studi dan nilai tes IQ sebelumnya kemudian digunakan sebagai acuan data dan perhitungan dengan menggunakan metode K-means yang hasilnya nanti digunakan sebagai acuan penjurusan siswa yang berdasarkan hasil nilai studi.

Berikut ini ada beberapa paper yang digunakan sebagai referensi pembelajaran yaitu sebagai berikut ini :

1. “ Aplikasi k-means untuk pengelompokan mahasiswa berdasarkan nilai body mass index (BMI) dan ukuran kerangka” Oleh Tedy Rismawan dan Sri Kusumadewi. Kasus ini menjelaskan masalah BMI dan ukuran kerangka seseorang. Apabila seseorang telah mengetahui nilai BMInya, orang tersebut maka dapat mengontrol berat badan sehingga dapat mencapai berat badan normal yang sesuai dengan tinggi badan. Sedangkan apabila orang tersebut mengetahui ukuran kerangka tubuhnya maka orang tersebut dapat mengontrol berat badannya agar dapat selalu berada dalam keadaan ideal [TED08].

Metode yang digunakan adalah metode k-means, hasil dari penelitian ini yaitu algoritma klasifikasi K-Means dapat digunakan untuk mengelompokkan mahasiswa berdasarkan status gizi dan ukuran kerangka.

Kekurangan dan kelebihan dari paper ini yaitu, diharapkan dapat dikembangkan proses clustering yang dapat digunakan untuk mahasiswa yang berjenis laki-laki maupun perempuan dan aplikasi hanya dikembangkan untuk clustering yang tidak hanya berdasarkan nilai BMI dan ukuran rangka, namun juga dapat berdasarkan status gizi, tekanan darah, dan sebagainya.

2. “Diagnosa potensi peserta didik”, Oleh Hamid Muhammad, Ph.D. kasus permasalahan pada paper ini yaitu, pembelajaran yang dilaksanakan selama ini bersifat massal, yang memberikan perlakuan dan layanan pendidikan yang sama kepada semua peserta didik. Hasil dari kesimpulan paper ini yaitu setiap siswa mempunyai tingkat kecakapan, kecerdasan, minat, bakat, dan kreativitasnya yang berbeda-beda. Strategi pelayanan pendidikan seperti ini memang tepat dalam konteks pemerataan kesempatan, akan tetapi kurang menunjang usaha mengoptimalkan pengembangan potensi peserta didik secara cepat [HAM04].

Kelebihan dan kekurangan dari paper ini yaitu, Strategi pelayanan pendidikan alternatif dalam manajemen pendidikan perlu dikembangkan untuk menghasilkan peserta didik yang unggul, melalui pemberian perhatian, perlakuan dan layanan pendidikan berdasarkan bakat minat dan kemampuannya. Agar pelayanan pendidikan yang selama ini diberikan kepada peserta didik mencapai sasaran yang optimal, maka pembelajaran harus diselaraskan dengan potensi peserta didik. Oleh karena itu guru perlu melakukan pelacakan potensi peserta didik.