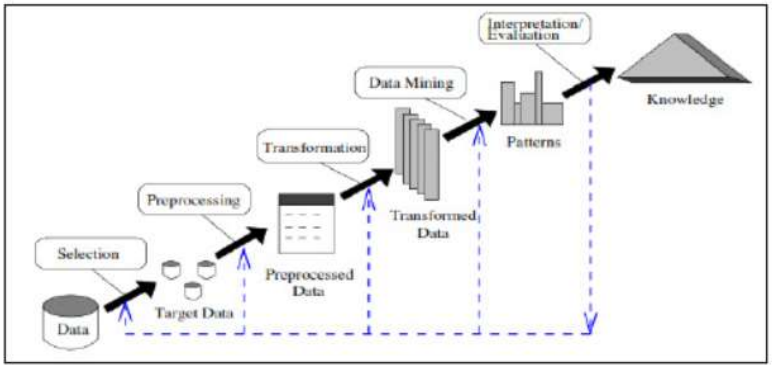
BAB II

TINJAUAN PUSTAKA

II.1. *Knowledge Discovery in Database* (KDD)

Istilah data *mining* dan *knowledge discovery indatabase (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah data *mining*(Marlina, Lina, *et al.*, 2018),.



Gambar II.1 ***Data Mining in Knowledge Discovery (KDD)***

Sumber : (Singh, Joshi and Mandhan, 2014)

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. Data *Selection*, penyeleksian data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing* / *Cleaning*, sebelum proses data mining dapat dilakukan, perlu proses pembersihan pada data yang menjadi fokus KDD. Proses

pembersihan yaitu membuang data ganda, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. *Transformation Coding* adalah transformasi data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung 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 terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data *mining* sangat bervariasi. Pemilihan metode atau algoritma yangtepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation / Evaluation*, pola informasi yang 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 interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya (Marlina, Putri, *et al.*, 2018).

II.2. Data Mining

Data mining adalah proses yang mengunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Salah satu teknik yang ada pada data mining adalah

klasifikasi (Harahap, 2015),(Kamila, Khairunnisa and Mustakim, 2019),(Katrina *et al.*, 2019).

Data mining adalah proses analitik yang dirancang untuk memeriksa sejumlah data yang besar dalam mencari suatu pengetahuan tersembunyi yang berharga dan konsisten [Florin G, 2011]. Tujuan dari data mining yaitu mencari trend atau pola yang diinginkan dalam database besar untuk membantu dalam pengambilan keputusan pada waktu yang akan datang (Pramesti, Furqon and Dewi, 2017),(Mujiasih, 2011),(Widyastuti *et al.*, 2019).

II.2.1. Pengelompokkan Data Mining

Data *Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Defiyanti and Jajuli, 2017),(Gustientiedina, Adiya and Desnelita, 2019) :

1. Deskripsi

   Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

   Estimasi hampir sama dengan klasifikasi, kecuali variabel target estmasi lebih ke arah numerik daripada ke arah kategori. Model

dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada penilaian berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model esrimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

   Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:

   - Prediksi harga beras dalam tiga bulan mendatang.

   - Prediksi presentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

4. Klasifikasi

   Di dalam klasifikasi terdapat target variabel kategori. Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Kemudian untuk menentukan pendapatan seorang pegawai, dipakai cara klasifikasi dalam data mining.

5. Pengklusteran

   Pengklusteran merupakan pengelompokkan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang mempunyai kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

6. Asosiasi.

   Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

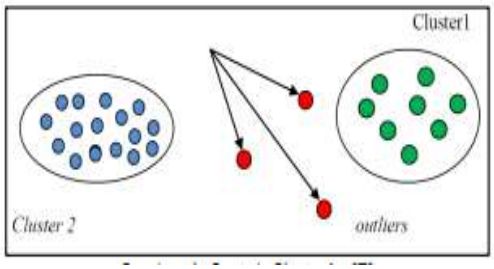
II.3. Clustering

Clustering merupakan suatu proses pengelompokkan record, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan klustering dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan pengelompokan pada proses clustering. *Clustering* sering dilakukan sebagai langkah awal dalam proses data *mining* (Pramesti, Furqon and Dewi, 2017),(Astuti, 2017),(Velmurugan and Santhanam, 2010).

Clustering merupakan suatu proses pengelompokan data, observasi, atau mengelompokan kelas yang memiliki kesamaan objek . Berbeda dengan proses klasifikasi, clustering tidak mempunyai target variable dalam melakukan. *Clustering sering dilakukan sebagai langkah awal dalam proses data mining* (Indriyani and Irfiani, 2019),(Anggara, Sujiani and Nasution, 2016). Clustering sering dilakukan sebagai langkah awal dalam proses data miningTerdapat

banyak algoritma klastering yang telah digunakan oleh peneliti sebelumnya seperti *K-Means, Improved K-Means, Fuzzy C-Means, DBSCAN, K-Medoids (PAM), CLARANS* dan *Fuzzy Substractive*. Setiap algortima memiliki kelebihan dan kekurangan masing-masing, namun prinsip algoritma sama, yaitu meelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok (Marlina, Putri, *et al.*, 2018),(Mustofa and Suasana, 2018).

Dalam data mining, usaha difokuskan pada metode - metode penemuan untuk cluster pada basis data berukuran besar secara efektif dan efisien. Beberapa kebutuhan klusterisasi dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda, mampu menangani dimensionalitas yang tinggi, menangani data yang mempunyai noise, dan dapat diterjemahkan dengan mudah. Dengan menggunakan klusterisasi, dapat mengidentifikasi daerah yang padat, menemukan pola - pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut - atribut data seperti pada gambar II.2.



Gambar II.2 Contoh Clustering

Sumber (Defiyanti and Jajuli, 2017)

II.4. Tunagrahita

Anak tunagrahita adalah individu yang secara signifikan memiliki intelegensi dibawah intelegensi normal. Menurut *American Asociation on Mental Deficiency* mendefinisikan Tunagrahita sebagai suatu kelainan yang fungsi intelektual umumnya di bawah rata- rata, yaitu IQ 84 ke bawah. Biasanya anak- anak tunagrahita akan mengalami kesulitan dalam "*Adaptive Behavior*" atau penyesuaian perilaku. Hal ini berarti anak tunagrahita tidak dapat mencapai kemandirian yang sesuai dengan ukuran (*standard*) kemandirian dan tanggung jawab sosial anak normal yang lainnya dan juga akan mengalami masalah dalam keterampilan akademik dan berkomunikasi dengan kelompok usia sebaya(Ardiyanto, 2014),(Ni Luh Gede Karang Widiastuti1), 2019),(Aminah, 2019). Definisi yang ditetapkan AAMD yang dikutip oleh Grossman, yang mengatakan artinya bahwa ketunagrahitaan mengacu pada sifat intelektual umum yang secara jelas dibawah rata-rata, bersama kekurangan dalam adaptasi tingkah laku dan berlangsung pada masa perkembangan. Dengan demikian dapat disimpulkan bahwa (Yosiani, 2014),(Louk and Sukoco, 2016):

a. Anak tunagrahita memiliki kecerdasan dibawah rata-rata sedemikian rupa dibandingkan dengan anak normal pada umumnya.
b. Adanya keterbatasan dalam perkembangan tingkah laku pada masa perkembangan.
c. Terlambat atau terbelakang dalam perkembangan mental dan sosial.
d. Mengalami kesulitan dalam mengingat apa yang dilihat, didengar sehingga menyebabkan kesulitan dalam berbicara dan berkomunikasi.

e. Mengalami masalah persepsi yang menyebabkan tunagrahita mengalami kesulitan dalam mengingat berbagai bentuk benda (visual perception) dan suara (audiotary perception).

f. Keterlambatan atau keterbelakangan mental yang dialami tunagrahita menyebabkan mereka tidak dapat berperilaku sesuai dengan usianya.

Contoh anak penderita tunagrahita pada sekolah SLB C Muzdalifah dapat dilihat pada gambar II. 3 .

Gambar II.3 Anak Penderita Tunagrahita

Sumber : SLB C Muzdalifah Medan (2020)

Pengklasifikasian anak tunagrahita penting dilakukan untuk mempermudah guru dalam menyusun program dan melaksanakan layanan pendidikan. Penting bagi Anda untuk memahami bahwa pada anak tunagrahita terdapat perbedaan individual yang variasinya sangat besar. Artinya, berada pada

level usia (usia kalender dan usia mental) yang hampir sama serta jenjang pendidikan yang sama, kenyataannya kemampuan individu berbeda satu dengan lainnya. Dengan demikian, sudah barang tentu diperlukan strategi dan program khusus yang disesuaikan dengan perbedaan individual tersebut. Pengklasifikasian ini pun bermacam-macam sesuai dengan disiplin ilmu maupun perubahan pandangan terhadap keberadaan anak tunagrahita.

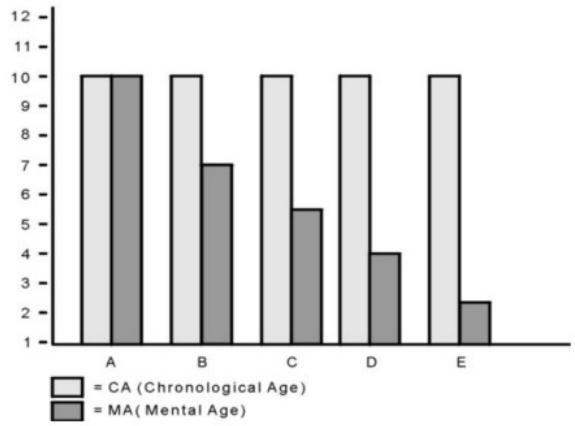
Klasifikasi anak tunagrahita yang telah lama dikenal adalah *debil, imbecile*, dan idiot, sedangkan klasifikasi yang dilakukan oleh kaum pendidik di Amerika adalah *educable mentally retarded* (mampu didik), *trainable mentally retarded* (mampu latih) dan *totally/custodial dependent* (mampu rawat). Pengelompokan yang telah disebutkan itu telah jarang digunakan karena terlampau mempertimbangkan kemampuan akademik seseorang. Klasifikasi yang digunakan sekarang adalah yang dikemukakan oleh AAMD sebagai berikut (Rochyadi, 2012)(Putri and Iswari, 2018),(Ismiyanti, 2015) :

1. *Mild mental retardation* (tunagrahita ringan)
2. *Moderate mental retardation* (tunagrahita sedang)
3. *Severe mental retardation* (tunagrahita berat)
4. *Profound mental retardation* (sangat berat)

Untuk memperjelas klasifikasi tersebut, perhatikan ilustrasi dan grafik berikut:

Ada lima orang anak berusia 10 tahun. Si A, IQ-nya 100 (normal); si B IQ-nya 70 -55; si C IQ-nya 55 - 40; si D IQ-nya 40 - 25; dan si E IQ-nya 25 ke bawah.

Untuk kebutuhan pendidikannya perlu ditentukan lebih dahulu umur kecerdasannya (mental age).



Gambar II.4 Grafik Klasifikasi Anak Berdasarkan Chronological Age dan Mental Age

Sumber (Rochyadi, 2012)

Dari grafik tersebut dapat dijelaskan sebagai berikut.

1. A berusia (chronological age) 10 tahun dan MA-nya 10 tahun.
2. B berusia 10 tahun dan MA-nya berkisar 7-5,5 tahun artinya ia dapat mempelajari materi pelajaran/ tugas anak normal usia 5,5 - 7 tahun.
3. C berusia 10 tahun dan MA-nya berkisar 5.5-4.0 tahun artinya ia dapat mempelajari materi pelajaran/ tugas anak normal usia 5,5-4.0 tahun.
4. D berusia 10 tahun dan MA-nya berkisar 4.0-2,5 tahun artinya ia dapat mempelajari materi pelajaran/ tugas anak normal 4,0-2,5 tahun.
5. E berusia 10 tahun dan MA-nya berkisar 2,5 tahun ke bawah artinya ia dapat mempelajari materi pelajaran/tugas anak normal usia 2,5 tahun ke bawah.

Klasifikasi yang digunakan di Indonesia saat ini sesuai dengan PP 72 Tahun 1991 adalah sebagai berikut :

1. Tunagrahita ringan,
2. Tunagrahita sedang,
3. Tunagrahita berat dan sangat berat.

II.5. Algoritma K-Medoids

K-Medoids atau *Partitioning AroundMedoids (PAM)* adalah algoritma *clustering* yang mirip dengan *K-Means.* Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* atau PAM menggunakan objek sebagai perwakilan *(medoid)* sebagai pusat cluster untuk setiap cluster, sedangkan *K-Means* menggunakan nilai rata-rata *(mean)* sebagai pusat cluster. Algoritma *KMedoids* memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma *K-Means* yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset (Pramesti, Furqon and Dewi, 2017),(Mustofa and Suasana, 2018), .

Algoritma *K-Medoids* hadir untuk mengatasi kelemahan Algoritma K-Means yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data. Langkah awal *K-Medoids* adalah mencari titik yang paling representatif (*medoids*) dalam sebuah dataset dengan menghitung jarak dalam kelompok dari semua kemungkinan kombinasi dari medoids sehingga jarak antar titik dalam suatu cluster kecil sedangkan jarak titik antar cluster besar. Algoritma *K-Medoids* memiliki

kelebihan untuk mengatasi kelemahan pada pada algoritma *K-Means* yang sensitive terhadap *noise* dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data (Silitonga *et al.*, 2019),(Rustam and Talita, 2015).

Langkah langkah algoritma *K-Medoids* (Marlina, Putri, *et al.*, 2018),(Sindi *et al.*, 2020),(Aryuni, Didik Madyatmadja and Miranda, 2018) . :

1. Inisialisasi pusat *cluster* sebanyak k (jumlah cluster)
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan II.1 :

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

   Dimana:

   d(x,y) = jarak antara data ke-i dan data ke-j

   xi1 = nilai atribut ke satu dari data ke-i

   yj1 = nilai atribut ke satu dari data ke-j

   n = jumlah atribut yang digunakan
3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika S < 0, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan cluster beserta anggota *cluster* masing-masing.