

BAB II

LANDASAN TEORI

2.1 IST (*Intelligenz Struktur Test*)

2.1.1 Tes Psikologi

Tes psikologi adalah alat ukur yang obyektif dan dibakukan atas sampel perilaku tertentu yang berfungsi untuk mengukur perbedaan-perbedaan antara individu-individu atau antara reaksi-reaksi individu yang sama dalam situasi yang berbeda (Anastasi & Urbina, 1997). Dalam kamus Psikologi, tes adalah satu perangkat pertanyaan yang sudah dibakukan, yang dikenakan pada seseorang dengan tujuan untuk mengukur perolehan atau bakat pada suatu bidang tertentu (Chaplin, 2006).

Sebuah tes psikologi pada dasarnya adalah alat ukur yang objektif dan dibakukan atas sampel perilaku tertentu (Anastasi & Urbina, 1997). Dalam psikologi, tes dapat diklasifikasikan menjadi empat, yaitu: Pertama, tes yang mengukur inteligensi umum (*general intelligence test*). Tes ini dirancang untuk mengukur kemampuan umum seseorang dalam suatu tugas. Kedua, tes yang mengukur kemampuan khusus atau tes bakat (*special ability test*). Tes ini digunakan untuk mengungkap kemampuan potensial subjek dalam bidang tertentu. Ketiga, tes yang mengukur prestasi (*achievement test*). Tes ini dimaksudkan untuk mengungkapkan kemampuan aktual sebagai hasil belajar. Keempat, tes yang mengungkap aspek kepribadian (*personality assesment*). Tes ini mengungkap karakteristik individual subjek dalam aspek non ability.

2.1.2 Teori Intelegensi

Louis Thurstone (1887-1955), yang menekankan pada aspek yang terbagi-bagi dari intelegensi. Thurstone menganggap bahwa intelegensi dapat dibagi menjadi sejumlah kemampuan primer. Menurut Thurstone, intelegensi umum yang dikemukakan oleh Spearman itu pada dasarnya terdiri dari 7 kemampuan

primer yang dapat dibedakan dengan jelas serta dapat digali melalui tes intelegensi, yaitu :

1. Pemahaman verbal (*verbal comprehension*), kemampuan memahami makna kata
2. Kefasihan menggunakan kata-kata (*word fluency*), kemampuan memikirkan kata secara tepat seperti penukaran huruf dalam kata, sehingga kata itu mempunyai pengertian lain atau memikirkan kata-kata yang bersajak
3. Kemampuan bilangan (*numerical ability*), kemampuan bekerja dengan angka dan melakukan perhitungan
4. Kemampuan ruang (*spatial factor*), kemampuan memvisualisasi hubungan bentuk ruang, seperti mengenal gambar yang sama yang disajikan dengan sudut pandang yang berbeda
5. Kemampuan mengingat (*memory*), kemampuan mengingat stimulus verbal
6. Kecepatan pengamatan (*perceptual speed*), kemampuan menangkap rincian visual secara cepat serta melihat persamaan dan perbedaan di antara obyek yang tergambar
7. Kemampuan penalaran (*reasoning*), kemampuan menemukan aturan umum berdasarkan contoh yang disajikan seperti menentukan bentuk keseluruhan rangkaian setelah disajikan sebagian dari rangkaian tersebut.

2.1.3 Gambaran Umum IST

Intelligenz Struktur Test (IST) adalah salah satu tes psikologis yang dapat mengukur inteligensi. IST disusun oleh Rudolf Amthauer di Jerman pada tahun 1973 (Van Der Ven, 1992) lalu diadaptasi oleh Universitas Padjajaran Bandung untuk penggunaannya di Indonesia. IST banyak digunakan di Indonesia untuk seleksi penempatan di dunia kerja maupun pendidikan karena IST yang terdiri dari sembilan sub tes tidak hanya menghasilkan skor IQ tetapi juga dapat

mengetahui minat dan bakat seseorang yang dapat dilihat dari skor setiap sub tes pada IST yang kemudian digunakan untuk membentuk profil kecerdasan praktis (dalam Hamidah & Hartini Nurul, 1985).

Tes IST adalah tes inteligensi yang terdiri dari 9 sub tes dengan jumlah itemnya 176 item dan merupakan tes kecepatan (*speed test*), Inteligensi dipandang sebagai suatu *gestalt* yang terdiri dari bagian-bagian yang saling berhubungan secara bermakna (struktur), sehingga dapat dipercaya bahwa struktur inteligensi tertentu akan cocok dengan tuntutan pekerjaan atau profesi tertentu pula. Oleh karena itu IST dapat menampilkan kemampuan khusus seseorang yang merupakan kekuatan dan kelemahan yang dimilikinya sesuai dengan perkembangan fungsi intelektualnya.

Tes IST digunakan untuk mengungkap kecerdasan sebagai kepandaian atau kemampuan untuk memecahkan periteman yang dihadapi. Inteligensi terdiri dari bagian-bagian saling berhubungan secara bermakna dan sebagai suatu *gestalt*. Struktur inteligensi tertentu menggambarkan pola bekerja yang tertentu yang akan cocok dengan tuntutan pekerjaan atau profesi tertentu. Ada pun tes yang digunakan akan meliputi beberapa interpretasi tiap-tiap subtes, yaitu (Polhaupessy, 1985):

1. Sub tes SE (*Satzergaenzung*). Dalam kamus Jerman (Adiwimarta dkk, 2008) *Satzer* berarti susunan kalimat. Item sub tes SE berupa 20 item pilihan ganda dan terdiri atas kalimat-kalimat. Pada setiap kalimat satu kata hilang dan disediakan lima kata pilihan sebagai penggantinya. Sub tes SE diuji pada pada subjek selama 6 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Pada sub tes ini berfungsi untuk mengetahui tentang potensi seseorang dalam:
 - a. Pembentukan keputusan, mengukur kemampuan seseorang dalam membuat kuputusan (dapatkah seseorang berprestasi)
 - b. Rasa realitas (menilai sesuatu yang mendekati reliabilitas)
 - c. *Common Sense* (memanfaatkan pengalaman masa lalu untuk mengatasi permasalahan) dapatkah seseorangberpikir secara mandiri)

- d. Berpikir konkrit praktis (masalah yang dihadapi sehari-hari)
 - e. Aspek yang diukur pada sub tes ini adalah *judgment* subjek.
2. Sub tes WA (*Wortauswahl*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Wortau* berartikan kata-kata. Item sub tes WA berupa 20 item pilihan ganda dan terdiri atas empat kata dan mencari kata kelima yang terdapat satu kesamaan. Sub tes WA diuji pada pada subjek selama 6 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Sub tes ini mengukur potensi seseorang dalam:
- a. Intelektual, rasa bahasa, kemampuan menghayati masalah bahasa, rasa empati.
 - b. Berpikir induktif dengan menggunakan bahasa, memahami pengertian
 - c. Pada remaja merupakan komponen intuisi
 - d. Pada orang dewasa merupakan motif sesuatu
3. Sub tes AN (*Analogien*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Analogien* berarti analogi, analisis. Item sub tes AN berupa 20 item pilihan ganda dan terdiri dari tiga kata dan mencari satu kata yang memiliki hubungan yang sama tersebut. Sub tes AN diuji pada pada subjek selama 7 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Sub tes ini berfungsi untuk mengukur potensi seseorang dalam:
- a. Fleksibilitas berpikir
 - b. Dapat berpikir logis atau menggunakan pikiran sebagai dasar berpikir (kedalaman hati)
 - c. Tidak suka bertindak berdasarkan kira-kira
 - d. Bila skor tinggi berarti mampu memahami hubungan antar masalah
 - e. Aspek yang diukur pada sub tes ini adalah proses berpikir yang mencakup analisis, *judgment* dan kesimpulan
4. Sub tes GE (*Gemeinsamkeiten*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Gemein* berarti hal yang umum. Item sub tes GE berupa 16

item menulis jawaban dan terdiri dari dua kata, dan mencari perkataan yang meliputi pengertian kedua kata. Sub tes GE diuji pada pada subjek selama 8 menit, untuk skoring jawaban benar dikali nilai 2, jawaban mendekati benar dikali nilai 1, sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes ini adalah untuk mengungkap kemampuan seseorang dalam :

- a. Kemampuan abstraksi, pembentukan pengertian
- b. Kemampuan untuk menyatakan pengertian dalam bahasa
- c. Membentuk suatu pengertian/mencari inti dari permasalahan
- d. Aspek yang diukur pada sub tes ini adalah kemampuan bernalar secara logis.

5. Sub tes ME (*Merkaufgaben*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Merkau* berarti memori. Item sub tes ME berupa 20 item pilihan ganda, subjek diminta untuk menghafal dan mengingat beberapa kata. Sub tes WU meminta subjek menghafal kata-kata selama 3 menit dan diuji selama 6 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes ini adalah untuk mengungkap potensi seseorang dalam:

- a. Mengukur kemampuan mengingat
- b. Indikasi konsentrasi yang menetap
- c. Konsentrasi dalam waktu yang relatif lama
- d. Sebagai tanda ketahanan
- e. Aspek yang diukur adalah memori atau ingatan yang berkaitan dengan perhatian dan konsentrasi.

6. Sub tes RA (*Rechenaufgaben*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Rechhe* berarti ilmu hitung. Item sub tes RA berupa 20 item pilihan ganda dan item menghitung. Sub tes RA diuji pada pada subjek selama 10 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes ini adalah mengukur kemampuan seseorang dalam :

- a. Cara berpikir praktis melalui hitungan
 - b. Kemampuan berhitung
 - c. Kemampuan menggunakan bilangan-bilangan secara praktis yang berhubungan dengan hitungan.
7. Sub tes ZR (*Zahlenreihen*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Zahlen* berarti menghitung. Item sub tes ZR berupa 20 item pilihan ganda dan terdiri deret angka yang harus dilanjutkan. Sub tes ZR diuji pada pada subjek selama 10 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes ini adalah untuk mengukur kemampuan seseorang dalam:
- a. Melihat momen-momen ritmis
 - b. Berpikir induktif dengan buangan secara teoritis (dengan angka)
 - c. Penggunaan bilangan secara agak teoritis
 - d. Berpikir teoritis dengan hitungan disertai dengan momen-momen ritmis.
8. Sub tes FA (*Formasuwahl*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Forma* berarti bentuk. Item sub tes FA berupa 20 item pilihan ganda dan terdiri atas item yang memperlihatkan sesuatu bentuk tertentu yang terpotong menjadi beberapa bagian. Sub tes FA diuji pada pada subjek selama 7 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes tersebut adalah untuk mengukur kemampuan seseorang dalam:
- a. Kemampuan membayangkan
 - b. Mengkonstruksi (sintesa dan analisa) tetapi ada momen konstantif
 - c. Berpikir secara konkrit yang menyeluruh
 - d. Memasukkan bagian pada suatu keseluruhan
 - e. Dapat menanggapi secara lebih variatif
 - f. Cara berpikir menyeluruh yang konkrit dalam momen-momen yang konstruktif.

9. Sub tes WU (*Wurfelaufgaben*). Dalam kamus Jerman (Adiwimarta dkk, 2008), *Wurfe* berarti kubus. Item sub tes WU berupa 20 item pilihan ganda dan terdiri atas lima kubus, yang mempunyai sisi yang berlainan. Dan setiap kubus memperlihatkan satu kedudukan yang berbeda. Sub tes WU diuji pada subjek selama 9 menit, untuk skoring jawaban benar dikali nilai 1 sedangkan jawaban salah dikali nilai 0. Fungsi dari sub tes ini adalah untuk mengukur kemampuan seseorang dalam:
 - a. Kemampuan membayangkan ruang tiga dimensi
 - b. Disertai dengan cara berpikir yang analitis

2.2 Pengelompokkan Jurusan Siswa

Struktur kurikulum SMA/MA meliputi substansi pembelajaran yang ditempuh dalam satu jenjang pendidikan selama tiga tahun mulai Kelas X sampai dengan Kelas XII. Struktur kurikulum disusun berdasarkan standar kompetensi lulusan dan standar kompetensi mata pelajaran. Pengorganisasian kelas-kelas pada SMA/MA dibagi ke dalam dua kelompok, yaitu kelas X merupakan program umum yang diikuti oleh seluruh peserta didik, dan kelas XI dan XII merupakan program penjurusan.

Penjurusan merupakan salah satu proses penempatan atau penyaluran dalam pemilihan program pengajaran para siswa di SMA/MA. Dalam penjurusan ini, siswa diberi kesempatan memilih jurusan yang paling cocok dengan karakteristik dirinya. Ketepatan dalam memilih jurusan dapat menentukan keberhasilan belajar siswa. Sebaliknya, kesempatan yang sangat baik bagi siswa akan hilang karena kekurangtepatan dalam menentukan jurusan. Menurut Permendiknas No. 22 Tahun 2016, Terdapat empat program jurusan yakni : Program Ilmu Pengetahuan Alam, Program Ilmu Pengetahuan Sosial, Program Bahasa, dan Program Keagamaan, khusus untuk MA.

Dalam kurikulum KTSP. Penjurusan di SMA dimulai pada akhir semester 2 kelas X. Selama di kelas X siswa hanya menerima program pengajaran umum,

sedangkan di kelas XI dan XII selain menerima program umum, siswa juga mendapatkan program pengajaran khusus sebagai pilihan IPA atau IPS.

Tujuan penjurusan antara lain adalah :

1. Mengelompokkan siswa sesuai dengan kecakapan, kemampuan, bakat dan minat yang relatif sama.
2. Membantu mempersiapkan siswa melanjutkan studi dan memilih dunia kerja.
3. Membantu memperkokoh keberhasilan dan kecocokan atas prestasi yang akan dicapai di waktu mendatang.

Ada beberapa faktor yang harus diperhatikan dalam penjurusan di SMA, yaitu :

1. Prestasi belajar

Kemampuan siswa dapat berwujud dalam kecakapan nyata dan kecakapan potensial. Kecakapan nyata dilihat antara lain dari prestasi belajar yang berbentuk skor atau nilai (hasil ulangan atau raport), sedangkan kecakapan potensial adalah salah satu kecakapan yang masih terpendam, yang dapat dilihat guru atau orang tua melalui alat non-tes seperti pengamatan, wawancara dan melihat prestasinya.

2. Minat siswa

Minat seseorang ditandai dengan rasa senang atau tidak senang, suka atau tidak suka. Minat timbul karena adanya informasi atau pengetahuan tentang suatu pekerjaan, benda atau situasi. Dalam hal ini kita selaku guru dan orang tua memberikan informasi dan pengetahuan yang benar dan tepat agar siswa mendapatkan gambaran yang jelas akan pilihannya.

3. Harapan orang tua

Berdasarkan pengalaman, ada orang tua yang memaksakan anaknya untuk masuk ke jurusan tertentu tetapi kemampuan anaknya tidak mendukung. Untuk itu, selaku pihak sekolah perlu mendengarkan atau memperhatikan keinginan/harapan orang tua terhadap anaknya, namun

seyogyanya juga perlu memberikan penjelasan tentang keadaan/kemampuan siswa sehubungan dengan pilihan tersebut.

4. Hasil psikotes

Tes psikologis adalah sebagai sarana untuk melengkapi hasil tes prestasi belajar, yaitu untuk mengukur kawasan-kawasan perilaku yang belum terungkap oleh tes prestasi belajar.

5. Daya tampung

Penjurusan disesuaikan dengan daya tampung sekolah, artinya berapa kelas sekolah tersebut menampung atau menerima program IPA, IPS, dan Bahasa ini tergantung kebijaksanaan atau ketentuan sekolah.

2.3 Data Mining

2.3.1 Pengertian Data Mining

Data mining merupakan suatu kegiatan yang meliputi pengumpulan, pemakaian dan historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dari data mining adalah pengelompokan clustering dimana data yang dikelompokkan belum mempunyai contoh kelompok. Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007).

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti database sistem, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, Image database, Signal processing (Han, 2006).

2.3.2 Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, memakai knowledge base. Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*Data selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil

dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*Data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*Pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

2.3.3 Teknik Data Mining

Beberapa teknik data mining antara lain (Bala., et al, 2012) :

1. Analisis asosiasi

Analisis asosiasi berupa penemuan aturan asosiasi yang menggambarkan kondisi atribut nilai yang sering terjadi bersamaan dalam sebuah satuan data tertentu. Analisis asosiasi secara luas digunakan untuk analisa data pasar dan transaksi

2. Klasifikasi dan Prediksi

Klasifikasi adalah pemrosesan untuk menemukan sebuah model yang menjelaskan dan memisahkan konsep atau kelas data, untuk kepentingan tertentu, yang bisa menggunakan pemodelan untuk memprediksi kelas objek yang labelnya tidak diketahui. Model yang didapat mungkin diwakili dalam berbagai format seperti aturan klasifikasi IF-THEN, pohon keputusan, formula matematika, atau jaringan syaraf tiruan pengklasifikasian bisa digunakan untuk memprediksi label kelas data objek data.

3. Analisis Clustering

Tidak seperti klasifikasi dan prediksi, yang menganalisa pelabelan objek data, clustering menganalisis objek data tanpa mengkonsultasikan label kelas yang dikenal. Secara umum label kelas bukan didapat dalam pengolahan data sederhana karena mereka tidak tahu bagaimana memulainya. Clustering dapat digunakan untuk me-generate label. Objek yang dicluster berdasarkan pada prinsip memaksimalkan persamaan dalam kelas dan meminimalkan kesamaan antar kelas. Sehingga cluster terhadap objek dibentuk sedemikian rupa sehingga objek dalam cluster mempunyai persamaan yang tinggi dalam perbandingan dengan objek lainnya, tapi sangat berlainan dengan objek dari cluster lain

4. Analisis Outlier

Sebuah database mungkin berisi objek data yang tidak sesuai dengan kebiasaan umumnya dari data yang disebut outlier. Analisa terhadap

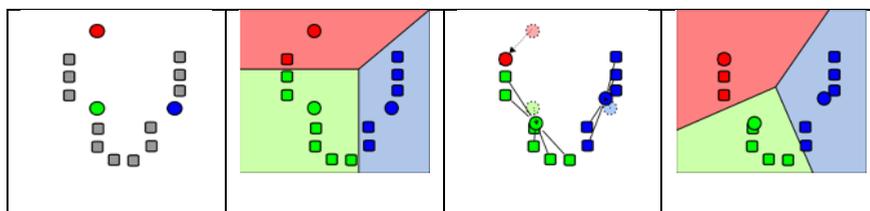
outlier mungkin membantu dalam pendeteksian kesalahan dan nilai-nilai abnormal.

2.4 Algoritma K-Means

Algoritma K-Means merupakan algoritma pengelompokan iteratif yang meakukan pertisi set data ke dalam sejumlah K cluster yang sudah ditetapkan di awal. Algoritma K-Means sederhana untuk dimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek (Prasetyo, 2014).

Secara historis, bentuk esensial K-Means ditemukan oleh sejumlah peneliti dari lintas disiplin ilmu. Yang paling berpengaruh adalah Lloyd (1982), Forgey (1965), Friedman dan Rubin (1967), dan McQueen (1967). Algoritma K-Means berkembang hingga menjadi konteks yang lebih besar sebagai algoritma *hill-climbing*, seperti yang disampaikan Gray dan Nuhoff (1998) dikutip oleh Prasetyo (2014)

K-Means mengelompokkan data berdasarkan jarak dari tiap dokumen ke pusat cluster (centroid) yang sudah ditentukan sebanyak K, dan mengelompokkan data-data ke pusat cluster yang terdekat.



Gambar 2.1 Ilustrasi proses K-Means Clustering

Rumus perhitungan centroid baru dari setiap cluster dicari dengan diformulasikan oleh persamaan berikut (Prasetyo, 2014) :

$$c_j = \frac{1}{N_k} \sum_{l=1}^{N_k} x_{jl} \dots \dots \dots (2.1)$$

c_j = centroid/rata-rata cluster ke-i untuk variable ke-j

N_k = jumlah data yang tergabung dalam anggota suatu cluster

Sedangkan untuk menemukan jarak dua data diformulasikan oleh persamaan *Euclidean Distance* berikut (Prasetyo, 2014) :

$$d_2(x, y) = \sqrt{\sum_{i=1}^r (x_i - y_i)^2} \dots \dots \dots (2.2)$$

x, y ; x_i, y_i = nilai fitur ke- i dari x dan y

r = jumlah fitur dalam vektor

Menurut Prasetyo (2014), Algoritma dari metode K-Means itu sendiri adalah sebagai berikut :

1. Inisialisasi : tentukan nilai K sebagai jumlah cluster yang diinginkan dan metrik ketidakmiripan (jarak) yang diinginkan. Jika perlu, tetapkan ambang batas perubahan fungsi objektif dan ambang batas perubahan posisi centroid.
2. Pilih K data dari set data X sebagai centroid.
3. Alokasikan semua data ke centroid terdekat dengan metrik jarak yang sudah ditetapkan (memperbarui cluster ID setiap data).
4. Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu (a) perubahan fungsi objektif sudah dibawah ambang batas yang diinginkan; atau (b) tidak ada data yang berpindah cluster; atau (c) perubahan posisi centroid sudah di bawah ambang batas yang ditetapkan.

2.5 *Davies-Bouldin Index*

Metrik Davies-Bouldin Index diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 Yang digunakan untuk mengevaluasi cluster. Validitas internal yang dilakukannya adalah seberapa baik clustering sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data (Prasetyo, 2014). *Sum of square within cluster (SSW)* sebagai metrik kohesi dalam sebuah cluster. Persamaan menghitung SSW adalah sebagai berikut (Prasetyo, 2014) :

$$SSW_i = \frac{1}{m} \sum_{j=1}^{m_i} d(x_j, c_i) \dots\dots\dots(2.3)$$

m_i = jumlah data yang berada dalam cluster ke- i

d = jarak ketidakmiripan

c_i = centroid cluster ke- i

Sementara matrik untuk separasi antara dua cluster, misalnya cluster i dan j , digunakan formula *sum of square between cluster* (SSB) dengan mengukur jarak centroid c_i dan c_j seperti pada persamaan berikut (Prasetyo, 2014) :

$$SSB_{i,j} = d(c_i, c_j) \dots\dots\dots(2.4)$$

Didefinisikan R_{ij} adalah ukuran rasio seberapa baik nilai perbandingan antara cluster ke- i dan cluster ke- j . Nilainya didapatkan dari komponen kohesi dan separasi. R_{ij} diformulasikan oleh persamaan berikut (Prasetyo, 2014) :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots\dots\dots(2.5)$$

Sifat-sifat yang dimiliki R_{ij} sebagai berikut :

1. $R_{i,j} \geq 0$
2. $R_{i,j} = R_{j,i}$
3. Jika $SSW_j < SSW_r$ dan $SSB_{i,j} = SSB_{i,r}$ maka $R_{i,j} > R_{i,r}$
4. Jika $SSW_j = SSW_r$ dan $SSB_{i,j} < SSB_{i,r}$ maka $R_{i,j} > R_{i,r}$

Nilai Davies-Bouldin Index (DBI) didapatkan dari persamaan berikut (Prasetyo, 2014) :

$$DBI = \frac{1}{K} \min_{i=1}^K \max_j (R_{i,j}) \dots\dots\dots(2.6)$$

K = jumlah cluster yang digunakan

Dari syarat-syarat perhitungan yang didefinisikan diatas, dapat diamati bahwa semakin kecil nilai SSW maka hasil clustering yang didapat juga lebih baik. Secara esensial, DBI menginginkan nilai sekecil (non-negatif ≥ 0) mungkin untuk menilai baiknya cluster yang didapat.

2.6 Penelitian Terkait

Penulis mengkaji hasil-hasil penelitian yang memiliki kesamaan topik dengan yang sedang diteliti oleh penulis. Adapun beberapa kajian yang berhubungan dengan topik yang sedang diteliti :

1. Alifah Umi Fathonah, 10.621.098, "*Pengelompokan Prestasi Akademik Mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik Menggunakan Metode K-Means*". Tahun 2014, Universitas Muhammadiyah Gresik. Kesimpulan dari penulisan ilmiah diatas adalah sebagai berikut : Pengelompokan dilakukan menggunakan atribut-atribut nilai matakuliah dari semester 1 sampai semester 6 yang berjumlah 41 matakuliah. Hasil uji coba sistem dengan evaluasi Davies-Bouldin Index yang paling kecil adalah 0.007149 dengan mengelompokkan mahasiswa menjadi 2 dan 3 klaster dengan prosentase 74.14% prestasi baik, 20.69% prestasi sedang dan 5.17% prestasi rendah.
2. Banatus Sa'adah, 09.622.034, "*Pengelompokkan Potensi Akademik Siswa RA Tarbiyatul Aulad Dengan Metode K-Means*". Tahun 2014, Universitas Muhammadiyah Gresik. Kesimpulan dari penulisan ilmiah diatas adalah sebagai berikut : Pengelompokkan potensi akademik siswa RA Tarbiyatul aulad dengan metode k-means merupakan pengelompokkan data nilai yang dipilih secara random, sehingga menemukan pusat cluster baru dan melakukan iterasi yang berulang-ulang sampai hasil matriks yang didapatkan clustering menjadi sama dan data tidak berubah.
3. M. Hamdi Asyrofi, 10.621.073, "*Pengelompokan Keterampilan Pemrograman Web Mahasiswa Teknik Informatika UMG Menggunakan Metode K-Means*". Tahun 2014, Universitas Muhammadiyah Gresik. Kesimpulan dari penulisan ilmiah diatas adalah sebagai berikut : Sistem klastering ketrampilan pemrograman WEB mahasiswa ini menggunakan metode K- Means. Pengelompokan

dilakukan dengan menggunakan atribut-atribut penilaian kompetensi dasar, kompetensi advance dan kompetensi PHP database. Hasil uji sistem dengan evaluasi klaster internal Davies-Bouldin Index yang paling bagus pada cluster 3 yakni 0,145. Dengan jumlah prosentase 63.53% mahasiswa kompetensi pemrograman WEB tinggi, 30.59 % mahasiswa kompetensi pemrograman WEB sedang dan 5.88 % mahasiswa kompetensi pemrograman WEB rendah.