

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pengertian Sekolah Menengah Atas (SMA)**

SMA merupakan jenjang pendidikan menengah yang mengutamakan penyiapan siswa untuk melanjutkan pendidikan yang lebih tinggi dengan pengkhususan. (Depdiknas, 2004: 112). Perwujudan pengkhususan tersebut berupa diselenggarakannya penjurusan di mulai di kelas XI (sebelas), yakni, penjurusan pada Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS), dan Bahasa yang mulai diterapkan pada siswa SMA kelas XI tahun pelajaran 2005/2006.

Penjurusan merupakan upaya strategis dalam memberikan fasilitas kepada siswa untuk menyalurkan bakat, minat, dan kemampuan yang dimilikinya yang dianggap paling potensial untuk dikembangkan secara optimal. Oleh karena itu, maka sekolah, guru, dan petugas Bimbingan Konseling (BK) harus mampu menempatkan ke dalam jurusan secara tepat.

Menempatkan siswa pada jurusan tertentu secara tepat berarti memberikan peluang kepada siswa untuk dapat berhasil pada masa yang akan datang. Hal ini sesuai dengan bunyi pasal 12 Undang-Undang Nomor 20 tahun 2003, bahwa Peserta didik mendapatkan pelayanan pendidikan sesuai dengan bakat, minat, dan kemampuannya. Untuk menghindari kemungkinan terjadi ketidaktepatan penjurusan siswa SMA dalam pilihan jurusan tertentu, perlu ada upaya antara lain: pengukuran dan penilaian keefektifan perencanaan penjurusan, keefektifan pelaksanaan penjurusan, keberhasilan siswa setelah penjurusan, serta kendala-kendala yang dihadapi dalam penjurusan.

#### **2.2 Kecerdasan Intelektual (IQ)**

Niai tes IQ adalah salah satu alat ukur kecerdasan seseorang. Kecerdasan ialah istilah umum yang digunakan untuk menjelaskan sifat pikiran yang mencakup sejumlah kemampuan, seperti kemampuan menalar, merencanakan,

memecahkan masalah, berfikir abstrak, memahami gagasan, menggunakan bahasa, dan belajar. Kecerdasan erat kaitannya dengan kemampuan kognitif yang dimiliki oleh individu (wikipedia, 2009).

*Intelligence Quotient* atau yang biasa disebut dengan IQ merupakan istilah dari pengelompokan kecerdasan manusia yang pertama kali diperkenalkan oleh Alferd Binet, ahli psikologi dari perancis pada awal abad ke 20. Kemudian Lewis Terman dari Universitas Stanford berusaha membakukan test IQ yang dikembangkan oleh Binet dengan mengembangkan norma populasi, sehingga selanjutnya tes IQ tersebut dikenal sebagai test Stanford Binet. Pada dasarnya kecerdasan intelektual merupakan kecerdasan tunggal dari setiap individu yang pada dasarnya beraturan dengan aspek kognitif dari setiap masing-masing individu tersebut. Tes Stanford Binet ini banyak digunakan untuk mengukur kecerdasan anak-anak sampai usia 13 tahun.

### **2.3 Data Mining**

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 system, 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).

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, Daniel T. 2005).

#### **1. Deskripsi**

Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

## 2. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning*.

## 3. Estimasi

Estimasi hampir sama dengan klasifikasi, perbedaannya adalah variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

## 4. Prediksi

Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, perbedaannya adalah hasil dari prediksi akan ada dimasa mendatang. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi.

## 5. Klastering

Klastering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Klastering menggunakan *unsupervised learning*.

## 6. Asosiasi

Tugas asosiasi atau sering disebut juga sebagai *market basket analysis* dalam data mining adalah menemukan relasi atau korelasi diantara himpunan item-item dan menemukan atribut yang muncul dalam satu waktu. Asosiasi menggunakan *unsupervised learning*. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* dan *confidence*.

Metode yang akan digunakan pada penelitian ini termasuk kedalam kelompok prediksi, karena menggunakan teknik klasifikasi yang hasilnya akan ada dimasa mendatang.

## 2.4 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Tahapan dari klasifikasi dalam data mining menurut (Han dan Kamber, 2006) terdiri dari :

### 1. Pembangunan Model

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

### 2. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut/class dari sebuah data baru yang atribut/classnya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

## 2.5 FK-NN (*Fuzzy K-Nearest Neighbor*)

Selain K-NN yang melakukan prediksi secara tegas pada uji berdasarkan perbandingan K tetangga terdekat, ada pendekatan lain yang dalam melakukan prediksi juga berdasarkan K tetangga terdekat, tetapi tidak secara tegas memprediksi kelas yang harus diikuti oleh data uji, pemberian label kelas data uji pada setiap kelas dengan memberikan nilai keanggotaan seperti halnya teori

himpunan *fuzzy*. Algoritma *Fuzzy K-Nearest Neighbor* (FK-NN) diperkenalkan oleh Keller *et al* (1985) dengan mengembangkan K-NN yang digabungkan dengan teori *fuzzy* dalam menyampaikan definisi pemberian label kelas pada data uji yang diprediksi.

Seperti halnya pada teori *fuzzy*, sebuah data mempunyai nilai keanggotaan pada setiap kelas, yang artinya sebuah data bisa dimiliki oleh kelas yang berbeda dengan nilai derajat keanggotaan dalam interval [0,1]. Teori himpunan *fuzzy* menggeneralisasi teori K-NN klasik dengan mendefinisikan nilai keanggotaan sebuah data pada masing-masing kelas. Formula yang digunakan (Liao, 2007):

$$u(x, c_i) = \frac{\sum_{k=1}^K u(x_k, c_i) * d(x, x_k)^{\frac{-2}{(m-1)}}}{\sum_{k=1}^K d(x, x_k)^{\frac{-2}{(m-1)}}} \dots \dots \dots (2.2)$$

Keterangan :

- u(x,c<sub>i</sub>)** = Nilai keanggotaan data x ke kelas c<sub>i</sub>.
- K** = Jumlah tetangga yang digunakan.
- u(x<sub>k</sub>,c<sub>i</sub>)** = Nilai keanggotaan data tetangga dalam K tetangga pada kelas c<sub>i</sub>, nilainya 1 jika data latih x<sub>k</sub> milik kelas c<sub>i</sub> atau 0 jika bukan milik kelas c<sub>i</sub>.
- d(x,x<sub>k</sub>)** = Jarak dari data x ke data x<sub>k</sub> dalam K tetangga terdekat.
- m** = Bobot pangkat yang besarnya >1.

Nilai keanggotaan suatu data pada kelas sangat dipengaruhi oleh jarak data itu ke tetangga terdekatnya. Semakin dekat ke tetangganya, semakin besar nilai keanggotaan data tersebut. Jarak tersebut diukur dengan N dimensi (fitur) data. Pengukuran jarak (ketidak miripan) dua data yang digunakan dalam FK-NN digeneralisasi dengan formula (Liao, 2007) :

$$D = \sqrt{(X1 + X2)^2 + (Y1 + Y2)^2} \dots \dots \dots (2.3)$$

Keterangan :

D = Jarak Euclidean

$X$  = Koordinat titik  $X$

$Y$  = Koordinat titik  $Y$

Meskipun FK-NN menggunakan nilai keanggotaan untuk menyatakan keanggotaan data pada setiap kelas, tetapi untuk memberikan keluaran akhir, FK-NN tetap harus memberikan kelas akhir hasil prediksi, untuk keperluan ini, FK-NN memilih kelas dengan nilai keanggotaan terbesar pada data tersebut.

### 2.5.1 Algoritma Fuzzy K-NN

Berikut ini adalah langkah-langkah yang perlu dilakukan dalam menerapkan metode Fuzzy K-NN dalam pengolahan data (Prasetyo, E.2012):

1. Normalisasikan data menggunakan nilai terbesar dan terkecil data pada setiap fitur.
2. Cari  $K$  tetangga terdekat untuk data uji  $x$  menggunakan persamaan (2.2).
3. Hitung nilai keanggotaan  $u(x, y_i)$  menggunakan persamaan (2.3) untuk setiap  $i$ , dimana  $1 \leq i \leq C$ .
4. Ambil nilai terbesar  $c = u(x, y_i)$  untuk semua  $1 \leq i \leq C$ ,  $C$  adalah jumlah kelas.
5. Berikan label kelas  $c$  ke data uji  $x$ .

### 2.5.2 Normalisasi Data Linear

Normalisasi data adalah proses penskalaan nilai pada atribut data yang cenderung memiliki nilai dengan rentang yang sangat bervariasi dan jenis data masukan berupa data *string*. Misalnya nilai jangkauan data yang besar akan memiliki pengaruh besar dalam menentukan jarak perhitungan dibandingkan nilai jangkauan data yang kecil (Larose,2005). Untuk itu dilakukan proses normalisasi data dengan menggunakan persamaan 2.1

$$X^* = \frac{(X - \min X)}{(\max X - \min X)} \dots \dots \dots (2.1)$$

Keterangan :

$X^*$  = nilai hasil normalisasi

$X$  = nilai  $x$  sebelum normalisasi,

**min** = nilai minimum dari fitur,  
**max** = nilai maksimum dari fitur,

### 2.5.3 Flowchart Perhitungan *K- Nearest Neighbor*



**Gambar 2.1** Flowchart *K-Nearest Neighbor*

dengan  $d$  adalah jarak antara titik pada data training  $x$  dan titik data testing  $y$  yang akan diklasifikasi, dimana  $x=x_1, x_2, \dots, x_i$  dan  $y=y_1, y_2, \dots, y_i$  dan  $I$  merepresentasikan nilai atribut serta  $n$  merupakan dimensi atribut (Han & Kamber, 2001). Sebagai ilustrasi, pada Tabel 2.1 berikut ini disajikan contoh penerapan rumus *Euclidean*, pada empat data klasifikasi kualitas baik dan tidak baik sebuah kertas tisu yang dinilai berdasarkan daya tahan

kertas tersebut dan fungsinya. Sebanyak tiga data yang sudah terklasifikasi yaitu data no 1,2, dan 3. masing-masing data dihitung jaraknya ke data no 4 untuk mendapatkan kelas yang sesuai bagi data no 4 maka k=1 klasifikasi kualitas baik atau tidak baik sebuah kertas tisu

**Tabel 2.1** Set data pada contoh metode FK-NN

| No | Fungsi | Daya Tahan | Klasifikasi |
|----|--------|------------|-------------|
| 1  | 7      | 7          | Tidak Baik  |
| 2  | 7      | 4          | Tidak Baik  |
| 3  | 3      | 4          | Baik        |
| 4  | 1      | 4          | ?           |

Berikut ini disajikan pula perhitungan yang dilakukan terhadap tiga data yang sudah terklasifikasi dengan data yang belum terklasifikasi pada Tabel di atas. Jarak data no satu ke data no empat:

$$d_{1,4} = \sqrt{(7 - 1)^2 + (7 - 4)^2} = \sqrt{6^2 + 3^2} = \sqrt{45} = 6.07$$

Jarak data no dua ke data no empat:

$$d_{2,4} = \sqrt{(7 - 1)^2 + (4 - 4)^2} = \sqrt{6^2 + 0^2} = \sqrt{36} = 6$$

Jarak data no tiga ke data no empat:

$$d_{3,4} = \sqrt{(3 - 1)^2 + (4 - 4)^2} = \sqrt{2^2 + 0^2} = \sqrt{4} = 2$$

Dari hasil perhitungan di atas diperoleh jarak antara data no tiga dan data no empat adalah jarak yang terdekat maka kelas data no empat adalah baik. Teknik ini akan diuji cobakan terhadap dataset yang belum terklasifikasi atau data yang belum dikenal, untuk menemukan kelas yang sesuai dengan berdasarkan pada data tetangga terdekatnya yang sudah terklasifikasi. Tingkat ketepatan klasifikasi terhadap data dari kedua algoritma yang digunakan menjadi titik fokus analisa dalam penelitian.

## 2.6 Evaluasi Sistem

Menurut (Weng dan Poon 2008) tujuan evaluasi percobaan pada klasifikasi yaitu untuk mengukur keefektifan apakah sistem mengklasifikasi

secara benar. Evaluasi biasanya membutuhkan sebuah matriks yang disebut berupa matriks *confusion*. Matriks *confusion* ditunjukkan pada tabel 2.2.

**Tabel 2.2** Matriks *confusion*

|            |         | Kelas Hasil Prediksi |                     |
|------------|---------|----------------------|---------------------|
|            |         | Positif              | Negatif             |
| Kelas Asli | Positif | True Positive (TP)   | False Negative (FN) |
|            | Negatif | False Positive (FP)  | True Negative (TN)  |

TP adalah jumlah dari kelas positif (kelas yang mempunyai jumlah data lebih sedikit) yang benar diklasifikasikan. FN adalah jumlah kelas positif yang salah diklasifikasikan ke dalam kelas negatif. FP adalah jumlah kelas negatif yang salah diklasifikasikan ke dalam kelas positif. TN adalah jumlah kelas negatif yang benar diklasifikasikan. Berdasarkan data yang didapatkan dari *confusion matrix*, akan dihitung (Weng dan Poon 2008) :

a. Akurasi dan Laju Error

Akurasi digunakan untuk mengukur prosentase pengenalan secara keseluruhan dan dihitung sebagai jumlah data uji yang dikenali dengan benar, dibagi dengan jumlah seluruh data uji. Berikut rumus akurasi dan laju *error* pada persamaan 2.4 dan rumus 2.5 *confusion matrix*.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \\
 &= \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \dots \dots \dots (2.4)
 \end{aligned}$$

$$\begin{aligned}
 \text{Laju error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{Jumlah prediksi yang dilakukan}} \\
 &= \frac{FN+FP}{TP+FN+FP+TN} \times 100\% \dots \dots \dots (2.5)
 \end{aligned}$$

## 2.7 Penelitian yang Terkait

Penelitian sebelumnya dilakukan oleh Nanik Yulianti (2012) dengan judul “*Aplikasi Sistem Pendukung Keputusan Penentuan Jurusan SMA Dengan Metode Topsis Dimadrasah Aliyah Negeri 2 Gresik*”. Tujuan dari penelitian tersebut adalah membuat sistem untuk mempermudah guru membantu mahasiswa dalam menentukan penjurusan siswa SMA. Dengan atribut-atribut yang digunakan nilai akademik siswa yaitu : matematika, fisika, kimia, biologi, sejarah, ekonomi, geografi, sosiologi dan nilai psikotes (tes IQ). Hasil dari penelitian tersebut, Sistem dapat merekomendasikan jurusan bagi siswa kelas X. Hasil akurasi dari penelitian tersebut adalah 49,73%.

Penelitian selanjutnya adalah tentang metode FK-NN (*Fuzzy K-Nearest Neighbor*) dalam penelitian yang berjudul “*Aplikasi Diagnosa Jenis Baja Berdasarkan Komposisi Kimia Dengan Menggunakan Metode FK-NN (Fuzzy K-Nearest Neighbor)*”, dibuat oleh Muhammad Sholikhuddin (Universitas Muhammadiyah Gresik, 2015). Tujuan dari penelitian tersebut adalah membuat sistem untuk mengetahui jenis baja sesuai dengan komposisi senyawa baja. Atribut-atribut yang digunakan dalam pemilihan baja yaitu senyawa : *Carbon, Silicon, Mangan, Fosfor, Sulfur, Chromium, Molybdenum, Wolfram, Vanadium dan Ferro rtc.* Hasil dari penelitian tersebut, Sistem dapat memprediksi data jenis baja. Hasil akurasi dari penelitian tersebut adalah 96,7% yang didapatkan dari 1-NN 4-NN dan 7-NN.

Maulidatul Hasanah (2015) dengan judul “*Klasifikasi Buah Naga Berbasis Tekstur Menggunakan Metode Fuzzy KNN*”. Dalam bidang pertanian, proses klasifikasi jenis buah dilakukan dengan indera penglihatan manusia, namun penilaian manusia dapat berbeda dari satu penilai dengan penilai lainnya sehingga diperlukan suatu standart penilaian yang sama. Pada tugas akhir ini diimplementasikan sebuah sistem yang akan mengklasifikasi *Buah Naga* berdasarkan tekstur kulit buah. Seleksi tekstur menggunakan analisis matriks kookurensi dan untuk proses klasifikasi menggunakan metode *fuzzy knn*. Dengan menggabungkan KNN dengan teori fuzzy akan dapat memberikan definisi pemberian label kelas pada data uji yang diprediksi.

Sistem dibuat dengan menggunakan program aplikasi Matlab R2008b. Pengujian dilakukan dengan menggunakan gambar buah Naga yang memiliki dimensi 448 x 336 *pixel* (dengan latar belakang yang sama) sebanyak 150 gambar. Dari hasil pengujian berdasarkan tekstur menggunakan metode *fuzzy knn* didapat hasil akurasi sebesar 90,6% untuk  $K = 6$ , untuk  $K = 8$  sebesar 92% dan 96% untuk  $K = 17$ .

Puspasari, M., Dewi, C., dan Rahman, M.A. 2013, Universitas Brawijaya. Penelitiannya yang berjudul : “*Prediksi Tingkat Resiko Penyakit Jantung Koroner (PJK) Menggunakan Metode Fuzzy K-Nearest Neighbor (FK-NN)*”. Pada penelitian ini digunakan metode gabungan Data Mining dengan *Logika Fuzzy*, yaitu *Fuzzy K-Nearest Neighbor (FK-NN)*. Dimana metode *Fuzzy K-Nearest Neighbor* dapat melakukan prediksi Penyakit Jantung Koroner sesuai dengan faktor-faktor yang ada. Dimana pada penelitian ini faktor PJK yang digunakan yaitu umur, HDL, trigliserida, LDL, kolesterol dan sistolik. FK-NN melakukan prediksi dengan mencari nilai tetangga terdekat kemudian menggunakan basis nilai keanggotaan data uji dari setiap kelas dengan mengambil nilai keanggotaan terbesar sebagai hasil akhir prediksi. Pengujian metode ini digunakan pada data latih PJK dengan jumlah berbeda. Dimana pada data latih dengan jumlah 40, 50, 60 dan 70 didapat akurasi tertinggi sebesar 66,67% pada data latih 60 dan 70. Semakin banyak data latih yang digunakan, maka kemungkinan semakin banyaknya jarak record yang mendekati kelas data prediksi.

Wisdarianto, A., Ridok, A., dan Rahman, M.A. 2013, Universitas Brawijaya. Penelitiannya yang berjudul “*Penerapan Metode Fuzzy K - Nearest Neighbor (FK-NN) untuk Pengklasifikasian Spam Email*”. Pada penelitian ini digunakan metode *Fuzzy K-NN* yang merupakan gabungan dari logika *Fuzzy* dan *K-NN*. Pada metode *Fuzzy K-NN* untuk klasifikasi email spam ini, email uji yang akan diklasifikasikan akan diberikan nilai keanggotaan untuk kelas spam dan kelas ham berdasarkan sejumlah  $k$  dokumen yang memiliki similarity tertinggi. Proses klasifikasi akan dilakukan dengan memilih nilai keanggotaan kelas tertinggi pada email dokumen uji tersebut. Pada penelitian ini dilakukan

pengujian 50 email uji terhadap penggunaan 50, 75,100, 125, 150, 175, hingga 200 data latih dan didapatkan hasil nilai f-measure paling baik yaitu 0,8979 pada penggunaan data latih 175 data latih untuk penggunaan nilai  $k=10$  dan  $k=15$ .

Sisca Lidhya Sari, Universitas Sumatera Utara. Penelitiannya yang berjudul ” *Implementasi Algoritma K-Nearest Neighbor Dan Metode Topsis ( Technique For Orders Preference By Similarity To Ideal Solution) Dalam Penentuan Mutu Beras Miskin (Studi Kasus: Bulog Aceh)* ” kriteria-kriteria yang digunakan dalam penelitian ini yaitu : Bebas Hama Penyakit, Butir Patah, Kadar Air, Bebas Bahan Kimia. Hasil dari penelitian tersebut, Sistem dapat dapat digunakan untuk menentukan mutu beras. Kesimpulan dari pengujian yang telah dilakukan, maka antara K-Nearest Neighbor dan TOPSIS terlihat algoritma K-Nearest Neighbor lebih sederhana proses perhitungannya terhadap data beras dibanding dengan Metode Topsis.

Yanita Selly Meristika, Achmad Ridhok, Lailil Muflikhah, Universitas Brawijaya. Penelitiannya yang berjudul “*Perbandingan K-nearest Neighbor Pada Diagnosis Penyakit Diabetes Mellitus*”. Pada penelitian ini Pengujian dilakukan terhadap 4 jumlah data latih yang berbeda yaitu 80, 130, 180, dan 230 dengan menggunakan jumlah data uji yang sama yaitu 50 data. Hasil pengujian yang dilakukan menunjukkan tingkat akurasi tertinggi terdapat pada FK-NN yakni mencapai 98%. Sedangkan K-NN akurasi tertingginya hanya mencapai 96%. Ini berarti Fuzzy K-Nearest Neighbor memberikan prediksi yang lebih baik dibandingkan K-Nearest Neighbor.