

BAB II

LANDASAN TEORI

2.1 Definisi Sistem

Sistem secara fisik adalah kumpulan dari elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (Gordon, 1991). Menurut (Jagianto, 2005) Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.2 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.

2.3 Data Mining

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. Klustering

Klustering 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 assosiatif 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 Normalisasi Data Linear

Normalisasi data linier adalah proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tertentu. Keuntungan dari metode ini adalah keseimbangan nilai perbandingan antara data saat sebelum dan sesudah nilai normalisasi. Kekurangannya adalah jika ada data baru metode ini akan memungkinkan terjebak pada out of bound error. Normalisasi data sangat di perlukan ketika data yang ada terlalu besar atau terlalu kecil sehingga pengguna kesulitan memahami informasi yang di maksud. Jika rentan nilai normalisasi yang di inginkan berada pada rentan [0,1] maka dapat juga menggunakan persamaan berikut :

$$\text{normalisasi } (X_{ik}) = \frac{X_{ik} - \min(X_k)}{\max(X_k) - \min(X_k)} \dots\dots\dots(2.2)$$

X= nilai yang akan dinormalisasi,

k = kolom kriteria,

i = baris data,

min = nilai terkecil dari semua data i,

max = nilai terbesar dari semua data i,

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):

$$u(x, c_i) = \frac{\sum_{k=1}^K u(x_k, c_i) * d(x, x_k)^{\frac{-2}{(m-1)}}}{\sum_{k=1}^K d(x, x_k)^{\frac{-2}{(m-1)}}} \dots\dots\dots(2.3)$$

Dengan:

$u(x, c_i)$ = Nilai keanggotaan data x ke kelas c_i .

K = Jumlah tetangga yang digunakan.

$u(x_k, c_i)$ = Nilai keanggotaan data tetangga dalam K tetangga pada kelas c_i , nilainya 1 jika data latih x_k milik kelas c_i atau 0 jika bukan milik kelas c_i .

$d(x, x_k)$ = Jarak dari data x ke data x_k 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) :

$$D = \sqrt{(X1 + X2)^2 + (Y1 + Y2)^2} \dots \dots \dots (2.4)$$

Dimana:

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.

2.6 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.

2.7 Penelitian Sebelumnya

Penelitian sebelumnya dilakukan oleh Masfufatul qouliyah (2016) dengan judul “ Pengelompokan potensi akademik siswa MTs AL-ibrahin menggunakan metode Self Organizing Map (SOM). Metode Self Organizing Map (SOM) dan salah satu algoritma yang digunakan dalam penelitian ini untuk menangani penilaian penilaian potensi siswa. Variabel yang digunakan dalam penilaian

potensi siswa di MTs AL-ibrahim adalah nilai mata pelajaran BHS.Indonesia, BHS.Ingggris, BHS.Arab, Matematika, IPA, IPS, Seni Budaya, PJ&Orkes, TIK, Quridits, Aqidah Akhlaq, Fiqih, SKI, Pkn, Nahwu, Aswaja, Shorof.

Berdasarkan dari hasil penelitian dan pembahasan yang dilakukan, SOM mampu mengklasifikasikan potensi siswa. Berdasarkan hasil pengujian di dapat akurasi dengan SOM adalah 71%.

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.

Penelitian selanjutnya dilakukan oleh 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 $K = 17$.

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.