

BAB II

LANDASAN TEORI

2.1 Konsep Pengelompokan (*Clustering*)

Pengelompokan atau *clustering* adalah melakukan pemisahan / pemecahan / segmentasi data ke dalam sejumlah kelompok (cluster) menurut karakteristik tertentu yang diinginkan. Dalam pekerjaan pengelompokan, label dari data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label sesuai keinginan (Prasetyo E., 2012). *Clustering* merupakan *unsupervised learning* yang membagi data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki masing-masing data tersebut. Awalnya karakteristik tiap kelompok tidak didefinisikan terlebih dahulu, namun karakteristik tersebut kemudian akan terlihat berdasarkan homogenitas karakter data dalam suatu kelompok yang akan berbeda dengan kelompok lainnya.

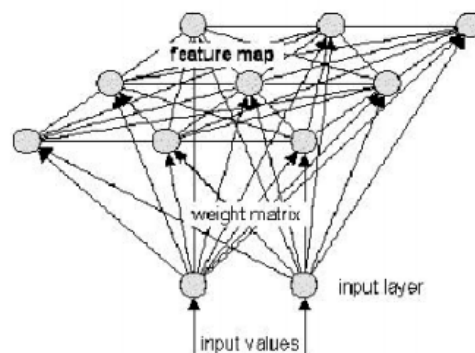
Analisis *cluster* merupakan suatu teknik analisa *multivariate* untuk mencari dan mengorganisir informasi tentang variabel sehingga secara relatif dapat dikelompokkan dalam kelompok yang homogen atau “cluster” dapat dibentuk. Cluster dibentuk dengan metode kedekatan yang secara internal harus homogen (anggota adalah serupa untuk satu sama lain) dan sangat secara eksternal tak sejenis (anggotanya tidak seperti anggota dari cluster yang lain). Analisis cluster dapat menerima suatu data masukan yang beragam. Ini biasanya disebut pengukuran “kesamaan”, dapat juga disebut “kedekatannya”, dan “kemiripannya”. Beberapa ahli merekomendasikan penggunaan standardisasi data, cluster dapat dihitung dalam skala yang berbeda dan standardisasi akan memberi pengukuran dengan menggunakan unit yang berbeda (Syaripudin U., 2013).

Tujuan pekerjaan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, biasanya proses pengelompokan dalam tujuan ini

hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok di mana sebuah data terletak di dalamnya (Prasetyo E., 2012).

2.2 *Self Organizing Maps (SOM)*

Self Organizing Maps (SOM) adalah salah satu metode dari jaringan syaraf tiruan yang diperkenalkan oleh Professor Teuvo Kohonen sekitar tahun 1980 an. SOM merupakan salah satu bentuk topologi dari *Unsupervised Artificial Neural Network (Unsupervised ANN)* dimana dalam proses pelatihannya tidak memerlukan pengawasan (target output). SOM digunakan untuk mengelompokkan (*clustering*) data berdasarkan karakteristik data. Arsitektur SOM dapat dilihat pada gambar 2.1



Gambar 2.1 : Arsitektur SOM

Pada SOM, suatu lapisan yang berisi neuron-neuron akan menyusun dirinya sendiri berdasarkan input nilai tertentu dalam suatu kelompok yang dikenal dengan istilah cluster. Selama proses penyusunan diri, cluster yang memiliki vektor bobot paling cocok dengan pola input (memiliki jarak paling dekat) akan terpilih sebagai pemenang. Neuron yang menjadi pemenang beserta neuron-

neuron tetangganya akan memperbaiki bobot – bobotnya (Lestari W., 2010). Berikut ini merupakan algoritma SOM (Prasetyo, E. 2012).

1. Inisialisasi bobot wij. Tentukan parameter topologi ketetanggaan. Tentukan parameter laju pembelajaran dan fungsi pembelajaran. Tentukan jumlah maksimal iterasi pelatihan.
2. Selama jumlah maksimal iterasi belum tercapai, lakukan langkah 3-7.
3. Untuk setiap data masukan X, lakukan langkah 4-6.
4. Untuk setiap neuron j, hitung menggunakan persamaan

$$D_j = \sum_i (w_{ij} - x_i)^2, i = 1, \dots, N \dots\dots\dots(2.1)$$
 dengan N adalah dimensi data (N)
5. Cari indeks dari sejumlah neuron, yaitu D_j , yang mempunyai nilai terkecil.
6. Untuk neuron j dan semua neuron yang menjadi tetangga J (yang sudah didefinisikan) dalam radius R dilakukan pembaruan bobot dengan persamaan

$$w_{ij}(\text{baru}) = w_{ij}(\text{lama}) + \eta (x_i - w_{ij}(\text{lama})) \dots\dots\dots(2.2)$$
7. Perbarui nilai laju pembelajaran

Pada algoritma di atas, parameter jarak untuk perbedaan atau kemiripan yang digunakan adalah Euclidean kuadrat (*square euclidean*). Hal ini dimaksudkan untuk mengurangi waktu komputasi dan menyederhanakan kinerja algoritma, tetapi harus dibayar dengan penggunaan memori yang lebih besar untuk alokasi nilai jarak yang biasanya besar. Nilai laju pembelajaran (η) yang digunakan menggunakan jangkauan nilai 0 sampai 1. Tetapi nilai ini akan terus diturunkan setiap kali ada kenaikan iterasi dengan sebuah fungsi. (Prasetyo, E. 2012).

2.3 Kelas Unggulan

Menurut Departemen Pendidikan dan Kebudayaan Republik Indonesia yang ditulis kembali oleh Agus Supriyono dalam tesisnya, kelas unggulan di Indonesia adalah suatu kelas yang dikembangkan untuk mencapai keunggulan dalam proses dan hasil pendidikan (Supriyono A., 2009).

Pada pasal 5 ayat 1 Undang-Undang nomor 20 tahun 2003 dinyatakan warga negara yang memiliki potensi kecerdasan dan bakat istimewa berhak untuk memperoleh pendidikan khusus. Selanjutnya pada pasal 12 ayat (b) disebutkan bahwa Setiap peserta didik pada setiap satuan pendidikan berhak mendapat pelayanan pendidikan sesuai dengan bakat, minat dan kemampuannya (Republik Indonesia, 2003).

Alasan pengelompokan peserta didik juga didasarkan atas realitas bahwa peserta didik secara terus-menerus bertumbuh dan berkembang. Pertumbuhan dan perkembangan peserta didik satu dengan yang lain berbeda. Agar perkembangan peserta didik yang cepat tidak mengganggu peserta didik yang lambat dan sebaliknya, maka dilakukanlah pengelompokan peserta didik (Soetopo H., 1982).

Dari pengertian tersebut dapat disimpulkan bahwa, kelas unggulan adalah kelas yang khusus dirancang untuk mengakomodir siswa yang memiliki potensi, bakat, kemampuan, kreativitas dan intelegensi yang lebih unggul daripada siswa yang lain untuk kemudian diberikan *treatment* pengajaran khusus guna meningkatkan potensinya sesuai dengan kurikulum yang dikembangkan.

2.4 Penjaringan Kelas Unggulan di MTs NU Trate Gresik

Kelas unggulan di MTs NU Trate Gresik dimulai pada jenjang kelas 8 dan 9. Pada masing – masing kelas unggulan tersebut dimasukkan siswa yang memiliki kemampuan akademis dan agama yang lebih unggul daripada siswa yang lain. Dengan mengelompokkan siswa tersebut, diharapkan dapat memberikan lingkungan belajar yang lebih seragam kondusif, kompetitif dan realtif sesuai dengan potensinya.

MTs NU Trate Gresik memiliki 15 pagu tiap tahunnya, dengan pembagian 5 pagu untuk masing – masing jenjang kelas. Sementara seleksi penjaringan kelas unggulan dilaksanakan setelah penerimaan raport semester genap tahun ajaran sebelumnya. Setelah dilakukan seleksi penjaringan, pada tahun ajaran 2017/2018 diperoleh 42 siswa unggulan pada jenjang kelas 8 dan 40 siswa unggulan pada kelas 9.

Sebagai sekolah berbasis agama Islam, MTs NU Trate Gresik mengadopsi dua kurikulum yakni kurikulum K13 dan KTSP yang berasal dari Kementerian Pendidikan dan Kebudayaan RI serta kurikulum agama dibawah naungan Kementerian Agama RI. Sesuai dengan kurikulum tersebut, maka kriteria penjarangan kelas unggulan di MTs NU Trate Gresik meliputi :

1. Nilai akademik, yakni nilai rata-rata raport mata pelajaran Ujian Nasional (B.Indonesia, B. Inggris, IPA dan Matematika) pada semester genap pada tahun ajaran 2016/2017.
2. Nilai agama, yaitu nilai rata-rata raport mata pelajaran agama (Al Qur'an Hadits, Aqidah Akhlak, Fiqih dan Sejarah Kebudayaan Islam) pada semester genap pada tahun ajaran 2016/2017.

2.5 *Confusion Matrix*

Confusion matrix adalah salah satu metode pengukuran kinerja (evaluasi) dari suatu model yang biasa digunakan dalam data mining.

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. *Accuracy* dalam klasifikasi adalah prosentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Han & Kamber, 2001).

Presisi atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar

Tabel 2.1 Model Confusion Matrix

Aktual	Classified as	
	+	-
+	True positives (A)	False negatives (B)
-	False positives (C)	True negatives (D)

Perhitungan *accuracy*, *precision*, dan *recall* dengan tabel *confusion matrix* dijabarkan masing – masing pada rumus sebagai berikut :

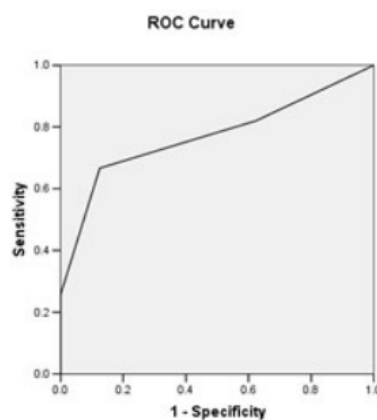
$$\text{Akurasi} = (A+D)/(A+B+C+D) \dots \dots \dots (2.3)$$

$$\text{Presisi} = A/(C+A) \dots \dots \dots (2.4)$$

$$\text{Recall} = A/(A+D) \dots \dots \dots (2.5)$$

Presisi dan Recall dapat diberi nilai dalam bentuk angka dengan menggunakan perhitungan persentase (1-100%) atau dengan menggunakan bilangan antara 0-1. Sistem rekomendasi akan dianggap baik jika nilai presisi dan recall bernilai tinggi. (Rosandy, 2016).

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan *false positive* sebagai garis horizontal dan *true positive* sebagai garis vertikal. AUC (*the area under curve*) dihitung untuk mengukur perbedaan performansi metode yang digunakan.



Gambar 2.2 : Kurva ROC

Tingkat diagnosa dari kurva ROC adalah sebagai berikut :

- a. Akurasi bernilai 0,90 – 1,00 = *excellent classification*
- b. Akurasi bernilai 0,80 – 0,90 = *good classification*
- c. Akurasi bernilai 0,70 – 0,80 = *fair classification*
- d. Akurasi bernilai 0,60 – 0,70 = *poor classification*
- e. Akurasi bernilai 0,50 – 0,60 = *failure*

2.6 Penelitian Sebelumnya

Dalam penelitian yang dilakukan oleh Syaripudin U.,dkk (2013) dengan judul Studi Komparatif Penerapan Metode Hierarchical, K-Means dan Self Organizing Maps (SOM) Clustering Pada Basis Data, mereka melakukan perbandingan penggunaan ketiga metode tersebut berdasarkan faktor data set

(besar dan kecil) serta faktor kluster (banyak dan sedikit). Dari hasil penelitian tersebut didapatkan dua kesimpulan, yang pertama bahwa algoritma SOM menghasilkan akurasi yang lebih baik dalam mengelompokkan objek ke dalam kelompok yang cocok dibandingkan dengan menggunakan algoritma k-means dan Hierarchical. Kesimpulan yang kedua dari penelitian tersebut menyatakan bahwa algoritma partisi (seperti k-means) direkomendasikan untuk data set berukuran besar sementara algoritma hierarchical clustering dan SOM direkomendasikan untuk data set berukuran kecil.

Penelitian yang dilakukan oleh Lestari W., (2010) berjudul Sistem Clustering Kecerdasan Majemuk Mahasiswa Menggunakan Algoritma Self Organizing Maps (SOM) bertujuan untuk menghasilkan sistem clustering pemetaan kecerdasan majemuk mahasiswa. Indikator kecerdasan majemuk yang dipakai berdasarkan *Gardner's Multiple Intelligence Scale*. Penelitian ini menghasilkan sistem yang dapat memetakan mahasiswa berdasarkan kemiripan kecerdasan majemuknya.

Penelitian selanjutnya dilakukan oleh Saraswati W., (2014) berjudul Clustering Menggunakan Self Organizing Maps (Studi Kasus : Data Perkembangan Anak di Kabupaten Bogor). Pada penelitian ini diimplementasikan clustering menggunakan SOM untuk data perkembangan anak dalam tiga kelompok rentang usia, yakni 2.5-3.4 tahun, 3.4-4.5 tahun dan 4.5-5.4 tahun. Variabel yang digunakan diantaranya kognitif, bahasa, motorik kasar dan motorik halus. Dari hasil percobaan ditemukan bahwa clustering data perkembangan anak rentang usia 2.5–3.4 tahun yang memiliki IDB minimal ialah ukuran cluster 3. Sementara itu, hasil percobaan pada rentang usia 3.5–4.4 ialah ukuran cluster 4, sedangkan pada rentang usia 4.5–5.4 tahun ialah ukuran cluster 3.

Selanjutnya, penelitian yang dilakukan oleh Ambarwati, dkk (2014) dengan judul Pengelompokan Berita Indonesia Berdasarkan Histogram Kata Menggunakan Self-Organizing Map, bertujuan untuk mengorganisir artikel berita secara lebih efisien dalam segi waktu dan biaya. Beberapa langkah yang mereka lakukan diantaranya preprocessing, tokenization, stopwords removing dan

stemming. Penelitian mereka membuahkan sebuah sistem yang dapat menampilkan hasil clustering dengan algoritma SOM dan menampilkan visualisasi dengan smoothed data histogram berupa island map dari artikel berita pada majalah tempo yang telah diproses.

Pada penelitian yang dilakukan oleh Harli E., dkk (2016) yang berjudul Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok. Mereka melakukan pengelompokan siswa berdasarkan 3 kluster yakni pintar, sedang dan cukup menggunakan metode SOM berdasarkan 6 variabel yang ditentukan oleh pihak sekolah, yakni prestasi (nilai rapor), kecerdasan (hasil psikotes), minat, jenis kelamin (variabel tambahan) dan nilai sikap (variabel tambahan). Hasil penelitian mereka menyimpulkan bahwa algoritma SOM dapat melakukan pengklasifikasian siswa secara merata pada setiap kelas yang ditentukan. Hasil tersebut berdasarkan penerimaan guru sebagai pengguna yang merasa hasil yang didapat sangat baik dan sesuai dengan penilaian nalar mereka sebagai manusia.

Sedangkan dalam konteks pengangkatan masalah yang hampir serupa, Wijilestari, dkk (2014) melakukan penelitian yang berjudul Klasifikasi Siswa Kelas Unggulan Menggunakan Fuzzy C-Means. Penelitian ini merupakan penelitian eksperimen, yakni melakukan pengujian tingkat akurasi algoritma FCM dalam seleksi penjarangan siswa kelas berdasarkan kriteria yang telah ditetapkan oleh sekolah. Kriteria penjarangannya meliputi nilai akademik, nilai potensi non akademik dan minat. Hasil penelitiannya menunjukkan bahwa algoritma FCM memiliki tingkat akurasi lebih tinggi (dengan rata-rata 86%) dibandingkan klasifikasi secara manual (dengan rata-rata 79%).