

BAB II

LANDASAN TEORI

2.1 Keberhasilan Studi

Kata “hasil” secara etimologi merupakan kata dasar dari kata “keberhasilan”. Di dalam Kamus Besar Bahasa Indonesia, kata “hasil” berarti sesuatu yang diadakan (dibuat, dijadikan, dan sebagainya) berdasarkan usaha. Sejalan dengan artian ini, kata “keberhasilan” mengandung makna perihal (keadaan) berhasil, tujuan dicapai. Berdasarkan pengertian demikian, “keberhasilan studi” dapat diartikan sebagai ketercapaian hasil dari tujuan studi.

UU Sisdiknas memberikan landasan hukum berkaitan dengan keberhasilan studi atau pendidikan nasional di Indonesia. UU Sisdiknas ini menekankan pentingnya keberhasilan studi bagi setiap individu yang sedang menjalani pendidikannya. Karena pentingnya ketercapaian keberhasilan studi ini, UU Sisdiknas menekankan bahwa ketercapaian keberhasilan studi bukan hanya tergantung individu yang sedang menjalankan studinya. Keberhasilan tersebut dapat dicapai juga karena peran serta warga negara, orangtua, masyarakat, dan pemerintah. Masing-masing hak dan kewajiban warga negara, orangtua, masyarakat, dan pemerintah yang berkaitan dengan pendidikan telah diatur. Lebih jauh, untuk mencapai keberhasilan studi tersebut, UU Sisdiknas juga memberikan landasan bagi sistem pendidikan nasional di Indonesia (Pemerintah RI, 2009).

Ada beberapa faktor yang dapat mempengaruhi keberhasilan studi. Secara garis besar, ada dua faktor yang dapat mempengaruhi, yaitu faktor internal dan faktor eksternal (Hakim, 2000). Pertama, faktor internal. Faktor internal ini merupakan faktor yang berasal dari dalam diri individu yang bersangkutan. Faktor ini berkaitan dengan faktor biologis dan faktor psikologis. Faktor biologis meliputi semua hal yang berhubungan dengan keadaan fisik. Keadaan fisik yang dimaksud adalah berkaitan dengan keadaan fisik yang normal atau tidak. Hal ini juga berkaitan dengan kondisi kesehatan fisik yang dimiliki. Selain faktor biologis, keberhasilan studi juga dapat dipengaruhi oleh faktor psikologis. Faktor psikologis ini merupakan segala sesuatu yang berkaitan dengan keadaan mental seseorang. Kondisi mental yang mantap dan stabil dapat menunjang seseorang untuk menggapai keberhasilan studinya. Berkaitan dengan faktor psikologis ini juga ada beberapa hal yang meliputinya, yaitu tingkat kecerdasan, kemauan, bakat, daya ingat, dan daya konsentrasi.

Kedua, Faktor eksternal. Faktor eksternal adalah faktor-faktor yang berada di luar diri individu yang sedang menjalankan studi. Kendati berada di luar, faktor eksternal ini dapat sangat mempengaruhi keberhasilan studi seseorang. Yang termasuk dalam faktor eksternal adalah keluarga, sekolah, masyarakat, dan waktu (kesempatan). Keluarga menjadi tempat pertama dan utama yang menentukan keberhasilan studi seseorang. Keadaan keluarga yang kondusif dengan keadaan ekonomi keluarga yang memadai akan mendukung seseorang dalam mencapai keberhasilan studinya. Demikian juga dengan kondisi masyarakat dan sekolah tempat terlaksananya pendidikan formal. Sekolah dengan mutu yang baik akan menunjang keberhasilan studi bagi mereka yang sedang mengenyam pendidikan

di dalamnya. Berkaitan dengan waktu, yang dimaksud adalah bagaimana individu yang sedang menjalankan studinya dapat mengatur waktu dengan baik. Di satu sisi bagaimana yang bersangkutan dapat memanfaatkan waktu untuk belajar, dan di sisi lain, bagaimana pula ia dapat mengisi waktu dengan kegiatan-kegiatan yang bersifat hiburan. Hendaknya ada keseimbangan yang proporsional dalam pemanfaatan waktu ini.

Sehubungan dengan keberhasilan studi dan faktor-faktor yang mempengaruhinya seperti yang telah disebutkan, seminari menengah sebagai tempat pendidikan calon imam Katolik juga memperhatikan hal-hal tersebut. Kriteria keberhasilan studi para seminaris sudah jelas berkaitan dengan ketercapaian kemampuan intelektual berkaitan dengan ilmu pengetahuan dan teknologi (*scientia et sapientia*), fisik dan mental yang sehat (*sanitas*), hidup berkomunitas (*socialitas*) yang baik, dan memiliki kesucian hidup spiritual (*sanctitas*). Semua kriteria ini harus terintegrasi secara utuh dalam pribadi seorang calon imam. Untuk menunjang ketercapaian tersebut, seminari harus juga memperhatikan faktor-faktor yang mempengaruhi keberhasilan studi para seminarisnya, baik itu faktor internal maupun faktor eksternalnya.

2.2 Data Mining Untuk Pendidikan (Educational Data Mining)

Educational Data Mining (EDM) atau *data mining* untuk pendidikan adalah disiplin ilmu yang mengembangkan penerapan *data mining* di dalam dunia pendidikan. Kenyataan yang ada saat ini, ada banyak sekolah, baik itu dari tingkat dasar sampai perguruan tinggi. Banyak sekolah yang tumbuh dan berkembang

diikuti banyaknya data yang ada di setiap sekolah tersebut. Data yang ada di dunia pendidikan ini, baik itu peraturan-peraturan atau kebijakan sistem pendidikan yang ada maupun data hasil pembelajaran setiap siswa, dapat berkembang seiring perjalanan waktu. Banyaknya data yang tersedia di dalam dunia pendidikan ini kiranya dapat digali guna mencari informasi yang berguna dan dapat diterapkan demi perkembangan dunia pendidikan itu sendiri.

Berkaitan dengan EDM, banyak sisi yang bisa diterapkan. Sisi yang dimaksud baik berkaitan dengan proses administratif di setiap sekolah, maupun sampai pada perkembangan peserta didik yang ada. Namun perlu disadari bahwa di dalam lingkup pendidikan, terdapat jenjang pendidikan, selain itu berkaitan dengan urutan waktu, dan konteks keberadaan sekolah juga dapat berpengaruh dan memiliki peran yang penting. Kendati banyak faktor yang perlu dipertimbangkan, namun bukan berarti penerapan *data mining* ini tidak dapat dilakukan.

Kendati EDM tergolong baru, dewasa ini sudah banyak tulisan dan penelitian yang dilakukan guna penerapan *data mining* di dunia pendidikan. Sudah banyak diskusi dan konferensi yang dilakukan guna pembahasan EDM ini. Konferensi internasional terakhir berkaitan dengan EDM, sampai saat tulisan ini dituliskan, diadakan pada 29 Juni - 2 Juli 2016 di Raleigh, Carolina Utara - USA. Ini adalah konferensi internasional kesembilan. Salah satu sumber informasi resmi berkaitan dengan EDM ini dapat dilihat di International Educational Data Mining Society (<http://www.educationaldatamining.org>).

Sehubungan dengan EDM, telah banyak bahasan dan eksperimen yang telah dilakukan. Ayala pada tahun 2013 telah mencoba untuk mengumpulkan 240 artikel atau jurnal yang berkaitan dengan EDM. Dari hasil penelitiannya, tulisan yang

berhubungan dengan EDM dapat digolongkan menjadi tiga bagian, yaitu berkaitan dengan kedisiplinan, tugas, dan metode yang ada di dunia pendidikan (Peña-Ayala, 2013). Selain itu, ada juga yang mencoba membangun sistem yang cukup kompleks untuk EDM. Sebagai contoh Gobert dan timnya (2015) mengimplementasikan *data mining* untuk membangun sistem penilaian yang kompleks berkaitan kinerja dan keterampilan yang ada di dunia pendidikan. Mereka menamakan sistem yang dibangun tersebut dengan Inq-ITS (Inquiryintelligent Tutoring System) (Gobert, Kim, Pedro, Kennedy, & Betts, 2015).

Sehubungan dengan EDM, karya yang dihasilkan tidak harus dalam skala yang besar. Ada juga yang mencoba untuk menyajikan kemampuan data mining dalam konteks lembaga pendidikan tinggi dan mencoba untuk menemukan pengetahuan eksplisit yang baru dengan menerapkan teknik *data mining*, hasilnya cukup memuaskan (Chalaris, Gritzalis, Maragoudakis, Sgouropoulou, & Tsolakidis, 2014). Tidak hanya berkaitan dengan lembaga, EDM juga dapat diterapkan untuk para siswa, misalnya berkaitan untuk karir yang cocok para peserta didik (Campagni, Merlini, Sprugnoli, & Verri, 2015). Selain memprediksi karir, sejalan dengan tulisan ini, Shahiri dan timnya (2015) memberikan gambaran tentang teknik *data mining* yang telah digunakan untuk mengklasifikasi kinerja siswa. Mereka berfokus pada bagaimana algoritma klasifikasi dapat digunakan untuk mengidentifikasi atribut yang paling penting dalam data siswa. Hasil yang diberikan cukup baik dan benar-benar bisa meningkatkan prestasi siswa dan kesuksesan lebih efektif dalam cara yang efisien dengan menggunakan teknik *data mining* untuk pendidikan (Shahiri, Husain, & Rashid, 2015).

Dari penelitian yang pernah dilakukan berkaitan dengan *data mining* untuk pendidikan, sejauh ini belum pernah ada yang mencoba untuk membuat penelitian di pendidikan khusus seperti seminari. Sejauh informasi yang dapat dihimpun, metode *data mining* yang pernah digunakan sejauh ini adalah menggunakan metode Decision Tree, Neural Network, Naïve Bayes, K-Nearest Neighbor, dan Support Vector Machine. Namun demikian masih sedikit penelitian yang menggunakan metode klasifikasi dengan metode optimalisasi untuk data hasil studi (Shahiri, Husain, & Rashid, 2015). Salah satu penelitian yang menggunakan metode optimasi dalam lingkungan pendidikan adalah dengan menggunakan Cuckoo Optimization dan Genetic Algorithm untuk menjawab masalah penjadualan matakuliah di universitas (Teoh, Wibowo, & Ngadiman, 2014)

Oleh sebab itu, tulisan ini mencoba untuk melihat lebih jauh akurasi yang dapat dihasilkan dengan menggunakan Jaringan Saraf Tiruan dan Jaringan Saraf Tiruan yang dioptimasi dengan Particle Swarm Optimization. Untuk sampai pada proses klasifikasi ini akan dilakukan seleksi fitur dengan menggunakan Multiclass F-score Feature Selection dan Principal Component Analysis untuk data hasil studi di seminari menengah.

2.3 Seleksi Fitur (Feature Selection)

Secara umum, metode klasifikasi menggunakan semua fitur yang terdapat di dalam data untuk membentuk sebuah model. Namun pada kenyataannya, belum tentu semua fitur tersebut relevan terhadap hasil klasifikasi. Jika hal ini terjadi pada data berdimensi besar, dapat mengakibatkan kinerja yang tidak efektif dan efisien. Oleh sebab itu, dibutuhkan seleksi fitur.

2.3.1 Multiclass F-score Feature Selection

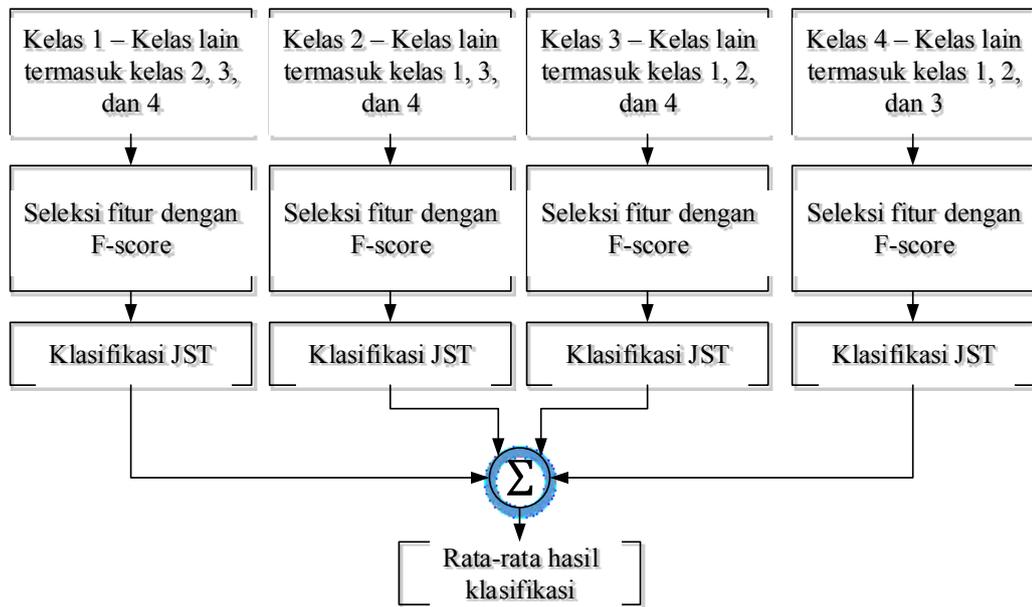
Sebelum melangkah ke proses klasifikasi, data dari seminari akan melalui proses *feature selection*. Multiclass F-score Feature Selection akan digunakan dalam tulisan ini. F-score adalah cara sederhana yang dapat mengukur diskriminan dua himpunan bilangan real. Vektor training $x_k, k = 1, 2 \dots, m$, jika jumlah kelas positif dan negatif misalnya n_+ and n_- , masing-masing, maka F-score dari sifat i^{th} didefinisikan sebagai berikut (Xu et al., 2008):

$$F_i = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (\bar{x}_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (\bar{x}_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (2.1)$$

Dari definisi di atas di mana $\bar{x}_i, \bar{x}_i^{(+)}, \bar{x}_i^{(-)}$ merupakan rata-rata dari fitur ke- i dari keseluruhan kumpulan data positif dan negatif *dataset*. Masing-masing $x_{k,i}^{(+)}$ adalah sifat ke- i dari contoh positif ke- k dan $x_{k,i}^{(-)}$ adalah sifat ke- i dari contoh negatif ke- k . Pembilangnya menunjukkan diskriminasi antara kumpulan bilangan positif dan negatif, dan penyebutnya menunjukkan yang ada di masing-masing dari dua himpunan tersebut. Semakin besar F-score, semakin besar pula kemungkinan sifat diskriminatif yang ada (Xu et al., 2008).

Untuk menerapkan F-score Feature Selection terhadap data yang memiliki banyak kelas (*multiclass*), prosedur tertentu harus dilakukan. Pencarian nilai F-score dilakukan dengan cara memisahkan data satu kelas dengan data lain selain data kelas tersebut yang telah digabungkan menjadi satu kelas tersendiri. Dengan demikian hanya ada dua kelas data. Setelah pemisahan ini dilakukan perhitungan F-score Feature Selection seperti persamaan 2.1 di atas. Kemudian dilakukan prosedur yang sama untuk setiap data kelas lainnya. Setelah pencarian fitur

selesai, masing-masing dari hasil seleksi fitur tadi masuk dalam metode klasifikasi. Pada tahap akhir dicari nilai rata-rata akurasi. Untuk lebih jelas dapat dilihat dalam gambar berikut ketika data yang digunakan contohnya memiliki empat kelas (Güneş, Polat, & Yosunkaya, 2010).



Gambar 2.1 Diagram Multiclass F-score Feature Selection

2.3.2 Principal Component Analysis (PCA)

Selain menggunakan F-score, data dari seminari juga akan diproses dengan metode Principal Component Analysis (PCA). PCA menghasilkan kombinasi linear dari variabel-variabel yang diperoleh dengan mereduksi variabel asli yang banyak. Di dalam proses mereduksi, diperoleh variabel yang lebih sedikit. Namun demikian variabel-variabel yang dihasilkan masih mengandung informasi yang termuat dalam data asli (Smith, 2002).

Dari sisi teknis, PCA dapat dimengerti sebagai suatu teknik mereduksi data multivariat (multivariable) yang mengubah (mentransformasi) suatu matriks data

asli menjadi kombinasi linier yang lebih sedikit, tetapi menyerap sebagian besar jumlah varian dari data asli tersebut. Atau secara singkat dapat dikatakan bahwa tujuan PCA adalah menjelaskan sebanyak mungkin jumlah varian data asli dengan menggunakan komponen utama (*principal component*) atau vektor sesedikit mungkin.

Untuk dapat menggunakan PCA dengan tepat, diperlukan pengetahuan dasar tentang perhitungan matematika. Hal ini dibutuhkan karena dalam perhitungan, PCA berkaitan dengan statistik dan aljabar matriks. Berikut ini adalah algoritma dalam penggunaan PCA (Smith, 2002):

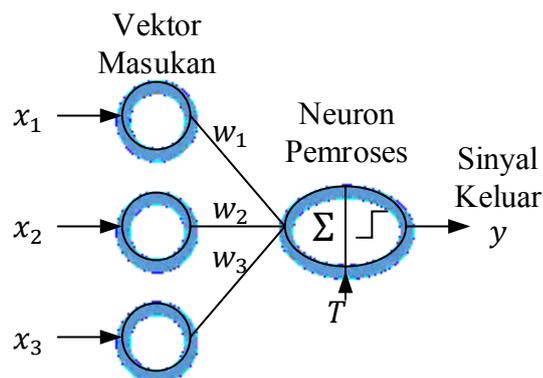
1. Matriks X adalah hasil pengurangan rata-rata dari setiap dimensi data pada matriks data.
2. Matriks C_x adalah *covariance matrix* dari matriks X .
3. Hitung *eigenvector* dan *eigenvalue* dari C_x .
4. Pilih *component* dan bentuk *vector feature* dan *principal component* dari *eigenvector* yang memiliki *eigenvalue* paling besar diambil.
5. Menurunkan data set yang baru.

2.1 Jaringan Saraf Tiruan (JST)

Jaringan Saraf Tiruan (JST) atau yang sering dikenal juga dengan *Artificial Neural Network* (ANN) adalah suatu konsep rekayasa pengetahuan dalam bidang kecerdasan buatan yang didesain dengan mengadopsi sistem saraf manusia di dalam otak. Bagian terkecil dari sistem saraf tersebut adalah sel saraf yang dikenal dengan unit dasar pemrosesan informasi yang dinamakan neuron. Disinyalir ada sekitar 10 miliar neuron dalam otak manusia dan 60 triliun koneksi antarneuron

tersebut. Dengan menggunakan secara simultan neuron-neuron tersebut, otak manusia dapat memproses informasi secara paralel dan cepat (Hagan, Demuth, Beale, & De Jesús, 1996).

Dalam sebuah neuron terdiri dari elemen-elemen badan sel (soma), sejumlah serat yang menyalurkan informasi ke neuron (dendrit), dan sebuah serat tunggal yang keluar dari neuron (akson). Setiap sinyal yang diterima oleh dendrit akan melewati sinaps untuk diteruskan ke neuron yang selanjutnya akan diproses di dalam soma. Setelah pemrosesan selesai akan dikeluarkan melalui akson untuk diproses kembali oleh neuron yang lain atau keluar sebagai sinyal akhir hasil proses di otak. Berdasarkan analogi kerja otak yang demikian JST juga terdiri dari sebuah unit pemrosesan yang disebut dengan neuron yang berisi penambah dan fungsi aktivitas sejumlah bobot. Secara umum, desain JST dapat dilihat dalam Gambar 2.2 di bawah ini (Prasetyo, E., 2012).

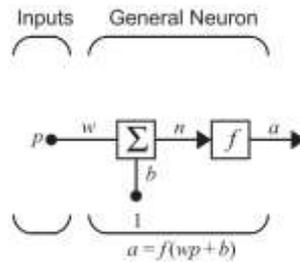


Gambar 2.2 Desain umum JST

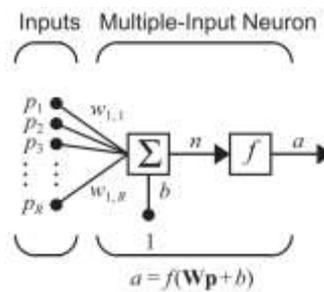
Serupa dengan otak manusia yang selalu belajar dari lingkungan sehingga dapat mengenali lingkungan dengan baik berdasarkan pengalaman yang sudah diperoleh, JST juga demikian. Dengan proses pembelajaran (*train*), JST di dalam *data mining* dianggap dapat menjadi model untuk melakukan prediksi suatu kelas data uji baru yang ditemukan. Dalam proses pelatihan, JST dapat menggunakan

algoritma tertentu seperti *Perceptron*, *Multilayer Perceptron*, *Backpropagation*, dan sebagainya (Prasetyo, E., 2012).

Ada berbagai tinjauan tentang JST. Dari sisi inputan, JST terdiri dari inputan tunggal dan inputan jamak. Arsitektur dari sisi inputan dapat dilihat seperti gambar berikut ini (Hagan et al., 1996).

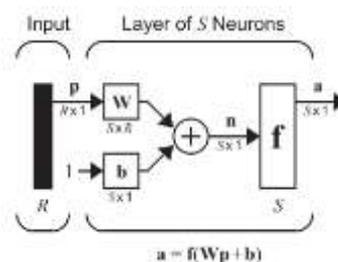


Gambar 2.3 JST dengan input tunggal

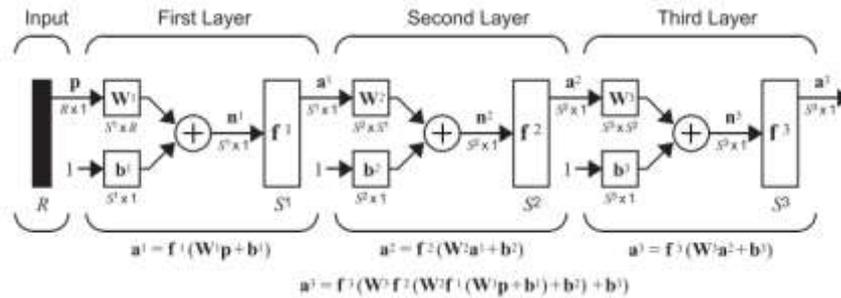


Gambar 2.4 JST dengan input jamak

Ditinjau dari jumlah lapisannya (*layer*), JST dapat dibagi ke dalam 2 jenis, yaitu JST dengan *layer* tunggal dan JST dengan *layer* jamak (*multilayer perceptron*). Berikut ini adalah gambaran umum JST dengan *layer* tunggal dan *layer* jamak (Hagan et al., 1996).



Gambar 2.5 JST dengan layer tunggal



Gambar 2.6 JST dengan tiga layer

Dari gambar di atas dapat dilihat bahwa JST, sama dengan metode lainnya memiliki input, baik tunggal maupun jamak, pemrosesan, dan *output* sebagai hasil yang diberikan. Sehubungan dengan hal ini, JST menggunakan perhitungan matematis dalam berbagai macam definisi untuk pemrosesannya. Di setiap inputan akan menerima bobot (w), begitu juga dengan setiap neuron akan memiliki nilai bias (b). Oleh karena itu apabila inputan adalah x , maka neuron (a) dapat didefinisikan demikian (Hagan et al., 1996):

$$a = w_1x_1 + w_2x_2 + \dots + w_nx_n + b \quad (2.2)$$

atau dalam bentuk matriks dapat juga didefinisikan sebagai berikut:

$$a = \mathbf{W}_x + b \quad (2.3)$$

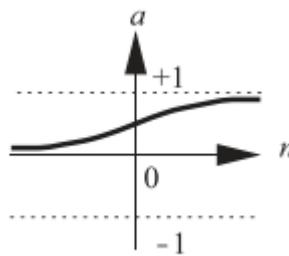
di mana \mathbf{W} adalah untuk neuron tunggal yang hanya memiliki satu baris.

Selanjutnya *output* (y) dapat didefinisikan sebagai berikut:

$$y = f(\mathbf{W}_x + b) \quad (2.4)$$

di mana f merupakan fungsi aktivitas. Ada berbagai fungsi aktifitas yang dapat digunakan baik itu fungsi linear maupun nonlinear. Fungsi aktivitas yang digunakan dalam tulisan ini adalah fungsi Sigmoid (*logsig*). Jika $a' = (\mathbf{W}_x + b)$ maka fungsi Sigmoid dapat didefinisikan sebagai berikut (Hagan et al., 1996):

$$a = \frac{1}{1+e^{-a'}} \quad (2.5)$$



Gambar 2.7 Grafik fungsi Sigmoid

Secara matematis, *output* dari proses JST ini untuk kelas binari yang adalah dasar dari klasifikasi *multiclass* dapat dituliskan sebagai berikut:

$$y = \begin{cases} 1, & \text{if } a > \theta \\ 0, & \text{otherwise} \end{cases} \quad (2.6)$$

2.2 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) didasarkan pada perilaku sebuah kawanan serangga, seperti semut, rayap, lebah atau burung. Algoritma PSO pada dasarnya meniru perilaku sosial organisme ini. Perilaku sosial yang dimaksud adalah terdiri dari tindakan individu dan pengaruh dari individu-individu lain dalam suatu kelompok. Kata partikel tertuju kepada, misalnya, seekor burung dalam kawanan burung. Setiap individu atau partikel berperilaku secara terdistribusi dengan cara menggunakan kecerdasannya sendiri. Hal ini juga dipengaruhi perilaku kelompoknya. Jika satu partikel, atau seekor burung menemukan jalan yang tepat atau pendek menuju ke sumber makanan, sisa kelompok yang lain juga akan dapat segera mengikuti jalan tersebut meskipun posisinya jauh dari kelompok (Santosa, 2009).

Dengan perilaku partikel dan kelompok organisme tersebut, disusun sebuah algoritma yang dapat berguna untuk proses optimasi. Algoritma atau metode

optimasi ini kemudian dikenal dengan PSO. Sebagai contoh adalah perilaku burung dalam kawanan burung:

1. Seekor burung tidak berada sangat dekat dengan burung lainnya.
2. Burung tersebut akan mengarahkan terbangnya ke arah rata-rata keseluruhan burung.
3. Burung tersebut juga akan memposisikan diri dengan rata-rata posisi burung yang lain dengan menjaga jarak sehingga burung dalam kawasan tersebut tidak terlalu jauh.

Hal ini menunjukkan bahwa perilaku kawanan burung akan didasarkan pada kombinasi dari tiga faktor, yaitu:

1. Kohesi, terbang bersama.
2. Separasi, menjaga jarak agar tidak terlalu dekat.
3. Alignment, penyesuaian untuk mengikuti arah bersama.

Berdasarkan perilaku yang demikian, PSO dikembangkan berdasarkan pada model berikut ini:

1. Ketika seekor burung mendekati target, makanan misalnya, lebih kurang akan secara cepat mengirimkan informasi kepada burung yang lain dalam kawasannya.
2. Burung yang lain akan mengikuti ke arah tujuan makanan tersebut, tetapi tidak secara langsung.
3. Masih ada komponen di dalam pikiran setiap burung, yaitu memorinya tentang sejarah yang sudah dilewati pada waktu sebelumnya.

Pemodelan ini disimulasikan dalam dimensi tertentu dengan sejumlah iterasi sehingga setiap iterasi, posisi partikel akan semakin mengarah kepada target yang

ingin dicapai. Hal ini dilakukan sampai mencapai iterasi maksimal atau bisa juga digunakan kriteria perhentian yang lain (Santosa, 2009).

Secara matematis untuk menggambarkan posisi dan kecepatan partikel pada suatu dimensi ruang tertentu adalah:

$$X_i(t) = x_{i1}(t), x_{i2}(t), \dots, x_{iN}(t) \quad (2.7)$$

$$V_i(t) = v_{i1}(t), v_{i2}(t), \dots, v_{iN}(t) \quad (2.8)$$

X : Posisi partikel

V : Kecepatan partikel

i : Indeks partikel

t : Iterasi ke- t

N : Ukuran dimensi ruang

Untuk menggambarkan mekanisme update status partikel dapat didefinisikan sebagai berikut:

$$V_i(t) = V_i(t-1) + c_1 r_1 (X_i^L - X_i(t-1)) + c_2 r_2 (X^G - X_i(t-1)) \quad (2.9)$$

$$X_i(t) = V_i(t) + X_i(t-1) \quad (2.10)$$

$X_i^L = x_{i1}^L, x_{i2}^L, \dots, x_{iN}^L$: Mempresentasikan *local best* dari partikel ke- i

$X_i^G = x_{i1}^G, x_{i2}^G, \dots, x_{iN}^G$: Mempresentasikan *global best* dari seluruh kawan

$c_1 c_2$: Konstanta bernilai positif (*learning factor*)

$r_1 r_2$: Bilangan *random* yang bernilai antara 0 sampai 1

Persamaan 2.7 digunakan untuk menghitung kecepatan partikel yang baru berdasarkan kecepatan sebelumnya, jarak antara posisi saat ini dengan posisi terbaik partikel (*local best*), dan jarak antara posisi saat ini dengan posisi terbaik kawan (*global best*). Pada langkah selanjutnya partikel terbang menuju posisi baru berdasarkan persamaan 2.8. Langkah ini dijalankan dalam sejumlah iterasi

sehingga mencapai kriteria tertentu sehingga akan didapatkan solusi yang terletak pada *global best* (Santosa, 2009).

2.7 Normalisasi Z-score

Z-score adalah metode sederhana untuk proses normalisasi data. Normalisasi ini sangat baik untuk mendekatkan kelompok nilai yang rentangnya jauh tanpa membuang informasi dari kelompok nilai tersebut. Z-score didefinisikan sebagai berikut (Abdullah, Rashid, Khan, & Musirin, 2015):

$$z = (x - \mu) / \sigma \quad (2.11)$$

x : Nilai individu

μ : Rata-rata suatu kelompok nilai

σ : Standar deviasi

2.8 Confusion Matrix

Untuk mengetahui seberapa jauh akurasi dari suatu hasil klasifikasi atau prediksi dibutuhkan evaluasi yang tepat. Dalam tulisan ini, ukuran yang akan digunakan untuk mengevaluasi hasil adalah dengan menggunakan Confusion Matrix. Confusion Matrix berisi informasi hasil aktual dan prediksi yang dihasilkan dari klasifikasi (Wilson, 2007).

Tabel 2.1 Tabel Confusion Matrix

	Aktual Positif	Aktual Negatif
Prediksi Positif	<i>TP</i>	<i>FP</i>
Prediksi Negatif	<i>FN</i>	<i>TN</i>

TP : *True Positif*, yaitu jumlah data positif yang diklasifikasikan benar oleh sistem.

FP : *False Positif*, yaitu jumlah data positif yang diklasifikasikan salah oleh sistem.

FN : *False Negative*, yaitu jumlah data negatif yang diklasifikasikan salah oleh sistem.

TN : *True Negative*, yaitu jumlah data negatif yang diklasifikasikan benar oleh sistem.

Dalam tulisan ini, evaluasi yang dilakukan dengan menggunakan tabel Confusion Matrix ini untuk mencari akurasi dari hasil prediksi yang diberikan oleh sistem. Adapun pencarian akurasi dapat didefinisikan sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (2.12)$$

