

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

*Backpropagation for Multi Label Learning* (BP-MLL) merupakan algoritma yang diturunkan dari algoritma jaringan syaraf tiruan yang sangat terkenal yaitu *Backpropagation* dengan cara menerapkan fungsi kesalahan baru yang menangkap karakteristik pembelajaran *multi label* (Zhang dan Zhou, 2006). Beberapa penelitian yang menggunakan algoritma BP-MLL dapat dilihat pada Tabel 2.1.

Tabel 2.1 Penelitian Terkait Backpropagation for Multi Label Learning

Penelitian	Task	Domain Data
Zhang dan Zhou (2006)	Memberi topik multi label pada artikel berita	Teks artikel
Zhang dan Zhou (2006)	Memberi anotasi gen multi label pada organisme jamur	Sel jamur
Zhang dkk (2018)	Membuat diagnosa dari <i>medical record</i>	Teks <i>medical record</i>
Usulan (2018)	Menghasilkan rekomendasi jurusan SMK berdasarkan psikotes bakat minat	Hasil psikotes bakat minat

Zhang dan Zhou (2006) menerapkan BP-MLL untuk memberi topik multi label pada artikel berita. Data latih yang digunakan pada penelitian tersebut diambil dari Reuters Collection dataset yang terdiri dari 21.578 artikel yang memiliki 135 label topik. Pada penelitian ini BP-MLL menghasilkan nilai akurasi sebesar 97%.

Zhang dan Zhou (2006) juga menerapkan BP-MLL untuk memberi anotasi gen pada sel jamur. Data yang digunakan pada penelitian tersebut diambil dari Yeast *Saccharomyces Cerevisiae* dataset dengan total 2417 gen dimana setiap gen

direpresentasikan dengan 103 variabel fitur. Pada penelitian ini BP-MLL menghasilkan nilai akurasi sebesar 79%.

Zhang, dkk (2018) menerapkan BP-MLL untuk menghasilkan rekomendasi diagnosa dari *medical record* berbahasa mandarin. Data latih yang digunakan pada penelitian tersebut berupa 10.000 *medical record* yang memiliki 737 label diagnosa. Pada penelitian ini BP-MLL menghasilkan nilai akurasi sebesar 74%.

Pada ketiga penelitian tersebut, terdapat suatu kesamaan yaitu seluruh data latih yang digunakan sudah memiliki multi label *ground truth*. Peneliti mengajukan menerapkan BP-MLL untuk menghasilkan rekomendasi jurusan SMK berdasarkan psikotes bakat minat. Berbeda dengan penelitian sebelumnya, data latih yang akan digunakan hanya memiliki keluaran 1 label, namun pada data baru yang akan diprediksi diharapkan dapat memberi keluaran multi label. Hal ini memungkinkan karena seperti yang disebutkan oleh Zhang & Zhou (2006) setiap pasang data harus memiliki paling tidak 1 label dan tidak boleh memiliki semua label.

## **2.2 Dasar Teori**

### **2.2.1 Penerimaan Peserta Didik SMK**

Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 14 Tahun 2018 tentang Penerimaan Peserta Didik Baru menyebutkan bahwa seleksi calon peserta didik baru kelas 10 SMK atau bentuk lain yang sederajat mempertimbangkan kriteria dengan urutan prioritas sesuai dengan daya tampung berdasarkan ketentuan rombongan belajar sebagai berikut:

- a. SHUN SMP atau bentuk lain yang sederajat; dan
- b. Prestasi di bidang akademik dan non-akademik yang diakui sekolah.

Selain mengikuti seleksi tersebut, sekolah dapat melakukan seleksi bakat dan minat sesuai dengan bidang keahlian/program keahlian/kompetensi keahlian yang dipilihnya dengan menggunakan kriteria yang ditetapkan sekolah dan institusi pasangan/asosiasi profesi.

Pada awal pendaftaran, calon siswa memilih 2 pilihan jurusan yang diinginkan dan tidak dapat diubah setelah diterima. Untuk itu, mengetahui jurusan yang cocok sebelum mendaftar merupakan hal yang krusial. Seleksi dilakukan dengan cara meranking nilai ujian nasional (Bahasa Indonesia, Bahasa Inggris, Matematika, Ilmu Pengetahuan Alam) dan prestasi calon siswa berdasarkan daya tampung per jurusan. Jika calon siswa tidak lolos seleksi pada pilihan jurusan pertama, maka calon siswa diseleksi di pilihan kedua. Apabila sekolah memilih untuk melakukan seleksi bakat dan minat, calon siswa yang tidak memenuhi kriteria calon jurusan dianggap tidak lolos seleksi.

### **2.2.2 Psikotes Bakat Minat**

Dalam bidang psikologi dikenal Psikotes Bakat Minat, yaitu serangkaian tes psikologi yang mampu menilai aspek-aspek utama kecerdasan, tipe kepribadian, dan bidang minat seseorang (Barret, 2009). Aspek kecerdasan memberi informasi tentang jenis-jenis kecerdasan yang relevan terhadap keterampilan dan karir yang berbeda. Aspek kecerdasan dapat memberi tahu kekuatan apa yang dimiliki seseorang dan menemukan cara untuk menggunakannya dan memberi tahu kelemahan seseorang namun tidak membiarkan dirinya untuk dibatasi oleh kelemahan tersebut. Aspek kepribadian dirancang untuk membantu memberi pemahaman bagaimana gaya seseorang dalam bekerja. Aspek minat memberi informasi kegiatan yang menarik bagi seseorang sehingga dapat membantu dalam memilih pekerjaan sehingga dapat menikmati apa yang ia lakukan setiap harinya. Psikotes Bakat Minat memberi informasi tentang seseorang di setiap aspek, namun hingga saat ini belum ada teori yang memetakan jurusan/bidang pekerjaan  $x$  harus memiliki variasi variabel seperti apa.

Pada masing-masing aspek psikotes bakat minat, terdapat banyak macam teori/alat tes yang dapat digunakan. Intelligence Structure Test, Rothwell Miller, dan Myers-Briggs adalah salah satu teori yang sangat populer digunakan oleh para psikolog di Indonesia. Intelligence Structure Test (IST) menjabarkan tingkat

intelegensi seseorang dalam 10 aspek (Furnham dan Crump, 2015). Daftar aspek tersebut dapat dilihat pada Tabel 2.2. Nilai aspek IQ adalah berupa angka antara 37 – 160, sedangkan pada aspek lainnya memiliki nilai dalam bentuk kelas nilai KS, K, C-, C, C+, CB, atau B.

Tabel 2.2 Daftar Aspek Penilaian Intelligence Structure Test

<b>Nama Variabel</b>	<b>Deskripsi</b>
IQ	Nilai Kecerdasan Umum
SE	Kemampuan common sense, pemahaman realita, berpikir praktis
WA	Kemampuan berpikir induktif, memahami pengertian bahasa
AN	Kemampuan fleksibilitas berpikir, daya mengkombinasikan & mendeteksi hubungan
GE	Kemampuan abstraksi verbal, berpikir logis dalam bentuk bahasa
RA	Kemampuan berpikir praktis dalam berhitung, kemampuan matematis
ZR	Kemampuan berpikir teoritis & induktif dengan angka-angka, kelincahan dalam berpikir
FA	Kemampuan mengamati, mengkonstruksi secara menyeluruh
WU	Kemampuan daya bayang ruang/tiga dimensi, beripikir analitis
ME	Kemampuan memori/daya ingat

Rothwell Miller membagi bidang minat studi/pekerjaan menjadi 12 bidang (Alias dan Abu Bakar, 2010). Daftar bidang tersebut dapat dilihat pada Tabel 2.3. Setiap bidang minat diberi nilai dalam bentuk kelas nilai kurang sekali, kurang, agak kurang, cukup, cukup kuat, kuat, atau kuat sekali.

Tabel 2.3 Daftar Bidang Minat Pada Teori Rothwell Miller

<b>Nama Variabel</b>	<b>Deskripsi</b>
Outdoor	Minat terhadap bidang yang berhubungan dengan kegiatan/aktifitas di luar ruangan
Mechanical	Minat terhadap bidang yang berhubungan dengan mesin, alat-alat, daya mekanik
Computational	Minat terhadap bidang/aktifitas yang berhubungan dengan angka
Scientific	Minat terhadap bidang yang terkait dengan analisa, penelitian, ilmu pengetahuan pada umumnya
Personal	Minat terhadap bidang yang mengharuskan kontak/berhubungan dengan orang lain, membujuk, mempengaruhi
Aesthetic	Minat terhadap bidang yang berhubungan dengan hal-hal yang bersifat seni, menciptakan sesuatu/kreasi
Library	Minat terhadap bidang yang berhubungan dengan buku-buku, kegiatan membaca, menulis/mengarang
Musical	Minat memainkan, mendengarkan alat musik, bernyanyi atau mendalami sesuatu terkait musik
Social	Minat terhadap bidang yang terkait dengan kepedulian atau kegiatan menolong, membimbing, melayani.
Administratif	Minat terhadap tugas-tugas yang menuntut ketepatan dan ketelitian
Practical	Minat terhadap pekerjaan yang praktis, karya pertukangan, tugas yang memerlukan ketrampilan
Medical	Minat terhadap pengobatan, penyembuhan penyakit, dan bidang biologis maupun medis pada umumnya

Myers-Briggs membagi tipe kepribadian seseorang dalam empat kategori dengan masing-masing memiliki 2 kelas (Furnham dan Crump, 2015). Daftar kategori tersebut dapat dilihat pada Tabel 2.4. Setiap kelas dalam setiap kategori diberi nilai berupa persentase dimana jika nilai kedua kelas ditambahkan akan berjumlah 100%.

Tabel 2.4 Daftar Kategori Pada Teori Myers-Briggs

Nama Kategori	Nama Variabel	Deskripsi
Orientasi Energi	Extraversion	Orientasi energi yang termotivasi melalui interaksi dengan orang lain, menyukai dunia luar, berorientasi pada tindakan & operasional, cenderung bertindak dahulu baru merefleksi
	Introversion	Orientasi energi yang lebih suka sendiri daripada berkumpul dengan banyak orang, cenderung menjadi pencetus ide, bekerja sendiri, penuh konsentrasi dan fokus
Pengelolaan Informasi	Sensing	Pengelolaan informasi yang realistis, menilai sesuatu berdasarkan fakta, berpedoman pada pengalaman/bukti, berorientasi pada masa kini
	Intuition	Pengelolaan informasi yang imajinatif, pemikiran konseptual, abstrak, lebih fokus pada gambaran besar daripada detail, lebih berorientasi pada kemungkinan masa depan daripada saat ini
Pengambilan Keputusan	Thinking	Pengambilan keputusan yang cenderung menggunakan logika, rasional, konsisten, lugas, dan objektif sehingga terkesan kaku dan keras kepala
	Feeling	Pengambilan keputusan yang melibatkan perasaan, empati dan nilai-nilai yang diyakini ketika mengambil keputusan, subjektif, akomodatif sehingga terkesan memihak, menginginkan harmoni
Orientasi Struktur	Judging	Orientasi struktur yang memiliki gaya hidup yang terstruktur, menentukan bagaimana seharusnya arah hidup, terkontrol, terencana, terorganisir
	Perceiving	Orientasi struktur yang memiliki gaya hidup fleksibel, lebih mudah beradaptasi dengan gaya hidup yang ada disekelilingnya

### 2.2.3 Sistem Pendukung Keputusan

Sistem pendukung keputusan adalah sistem yang dirancang untuk mendukung pengambilan keputusan pada masalah-masalah yang tidak terstruktur (Turban dkk, 2004). Sistem pendukung keputusan merupakan mekanisme untuk interaksi antara pengguna dan komponen-komponen lainnya. Komponen-komponen sistem pendukung keputusan tersebut adalah modul manajemen data, modul manajemen model, dan *user interface*.

Modul manajemen data berupa sistem manajemen basis data yang terdiri dari data-data yang diperoleh dari sumber internal maupun sumber eksternal. Modul manajemen model merupakan model kuantitatif yang dapat memberikan sistem kemampuan untuk menganalisis masalah dan menemukan solusinya. *User interface* adalah komponen yang menyediakan komunikasi antara pengguna dengan sistem pendukung keputusan.

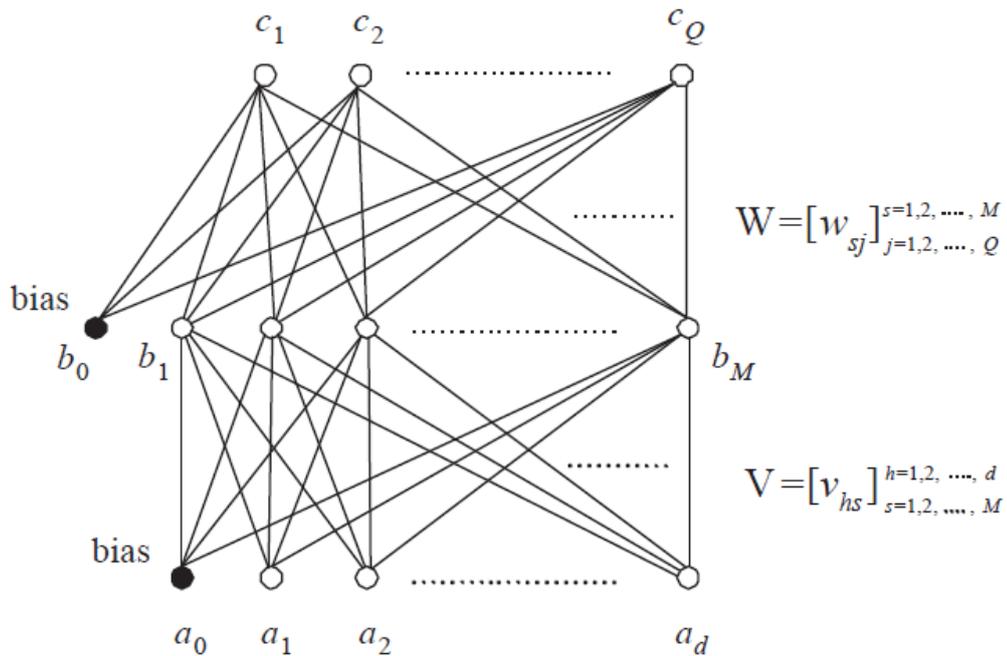
### 2.2.4 Backpropagation For Multi Label Learning

Backpropagation for Multi Label Learning (BP-MLL) merupakan algoritma turunan dari algoritma jaringan syaraf tiruan *multi-layer feed-forward* tradisional yang terkenal yaitu Backpropagation yang kemudian diadaptasi untuk belajar dari data latih multi label. Perbedaannya, Backpropagation tradisional hanya berfokus pada diskriminasi label individu, yaitu apakah label  $a$  milik *instance*  $Y$  atau bukan, sedangkan BP-MLL mempertimbangkan korelasi antar label yaitu yang label seharusnya memiliki ranking lebih tinggi daripada yang bukan label.

Untuk mengadaptasi Backpropagation tradisional dari penanganan data label tunggal menjadi data multi-label membutuhkan dua kunci (Zhang dan Zhou, 2006). Kunci pertama adalah mendesain beberapa fungsi kesalahan spesifik selain fungsi sum-of-squares sederhana untuk menangkap karakteristik pembelajaran multi-label. Kedua, beberapa revisi harus dibuat dari algoritma pembelajaran klasik untuk meminimalkan fungsi kesalahan yang baru dirancang.

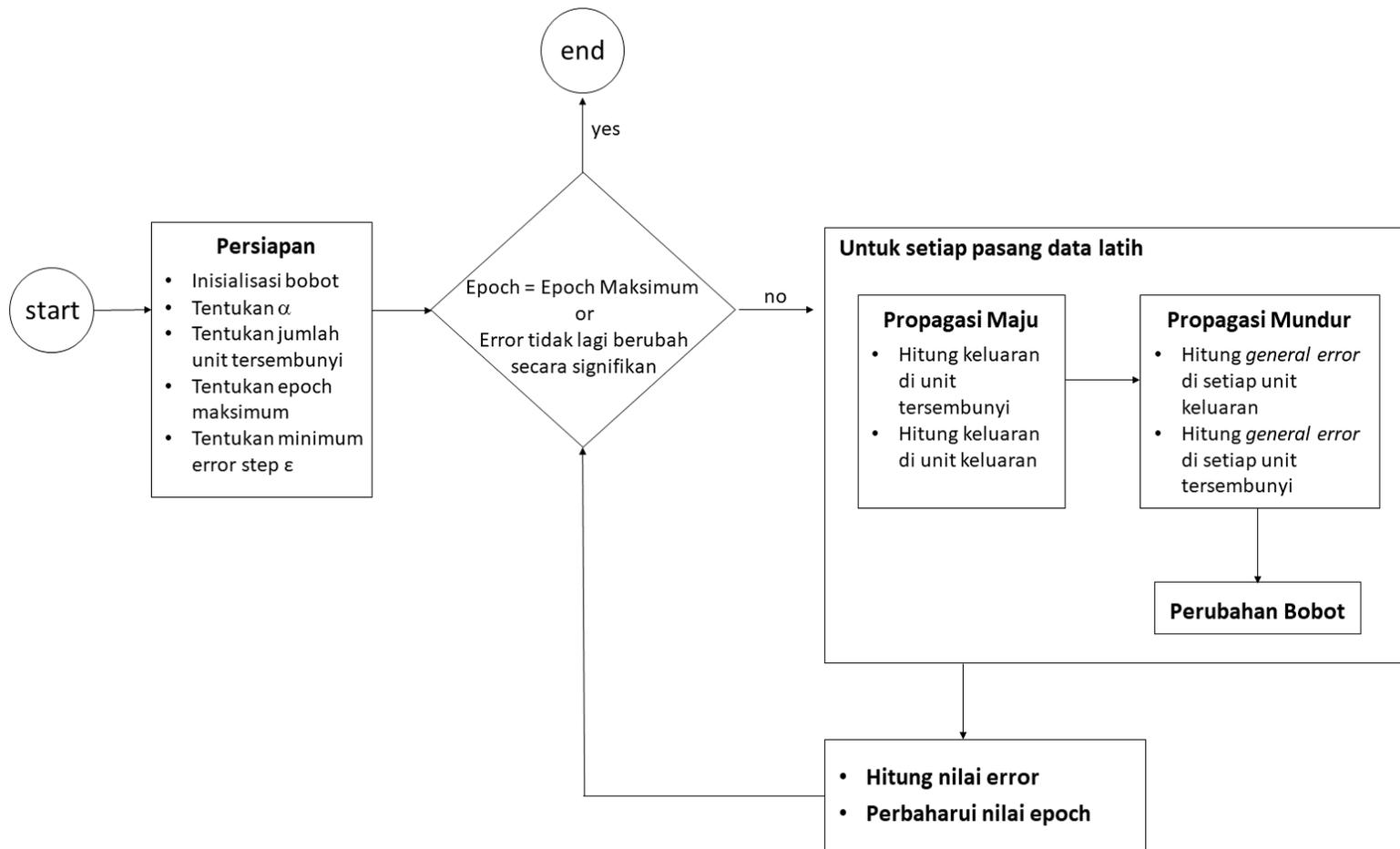
### 2.2.4.1 Arsitektur Jaringan Backpropagation For Multi Label Learning

Arsitektur jaringan BP-MLL terdiri dari satu atau lebih unit masukan ditambah satu unit bias, satu layer tersembunyi yang terdiri dari satu atau lebih unit ditambah satu unit bias, serta satu atau lebih unit keluaran.



Gambar 2.1 Arsitektur Jaringan BP-MLL

Gambar 2.1 adalah contoh ilustrasi arsitektur jaringan BP-MLL. Jaringan BP-MLL memiliki  $d$  unit masukan,  $Q$  unit keluaran yang setiap unitnya merupakan anggota label yang dimungkinkan, serta satu layer tersembunyi dengan  $M$  unit tersembunyi. Setiap unit masukan terhubung dengan semua unit tersembunyi dengan bobot  $V = [v_{hs}]$  ( $1 \leq h \leq d$ ,  $1 \leq s \leq M$ ) dan setiap unit tersembunyi terhubung dengan semua unit keluaran dengan bobot  $W = [w_{sj}]$  ( $1 \leq s \leq M$ ,  $1 \leq j \leq Q$ ). Sebuah ekstra unit masukan  $a_0$  yang disebut bias memiliki nilai tetap yaitu 1, serta pada unit tersembunyi juga terdapat bias  $b_0$  dan memiliki nilai tetap yaitu 1.



Gambar 2.2 Diagram Alur Algoritma Pelatihan BP-MLL

### 2.2.4.2 Algoritma Jaringan Backpropagation For Multi Label Learning

Algoritma BP-MLL menggunakan konsep *supervised learning*. Konsep *supervised learning* menggunakan data latih, yaitu data yang diambil dari himpunan model data yang diizinkan atau dimungkinkan, sehingga pada dasarnya data latih dapat diambil dari data yang terjadi di dunia nyata. Sepasang data pelatihan terdiri dari  $m$  instan multi label dimana  $x_i$  merupakan input dari jaringan dan  $y$  merupakan output multi-label.  $Y_i \subseteq y$  adalah set label yang dimiliki oleh sebuah instance, dan  $\bar{Y}_i$  merupakan set komplementer dari  $Y_i$ . Alur algoritma pelatihan BP-MLL dapat dilihat pada Gambar 2.2.

Untuk setiap pasang data latih  $x_i$  dan set label yang dimiliki ( $Y_i$ ), *output* sebenarnya dari unit keluaran ke- $j$  dapat dihitung menggunakan persamaan 2.1.

$$c_j = f(c\_net_j + \theta_j) \dots\dots\dots (2.1)$$

$\theta_j$  adalah bias dari unit keluaran ke- $j$ , dan  $f(x)$  adalah fungsi aktivasi yang berupa fungsi “tanh” yang dijabarkan pada persamaan 2.2.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \dots\dots\dots (2.2)$$

$c\_net_j$  adalah *input* dari unit keluaran ke- $j$  yang dapat dihitung menggunakan persamaan 2.3.

$$c\_net_j = \sum_{s=1}^M b_s W_{sj} \dots\dots\dots (2.3)$$

$W_{sj}$  adalah bobot yang menghubungkan unit tersembunyi ke- $s$  dan unit keluaran ke- $j$ .  $M$  adalah banyaknya unit tersembunyi.  $b_s$  adalah *ouput* dari unit tersembunyi ke- $s$  yang dapat dihitung menggunakan persamaan 2.4.

$$b_s = f(b\_net_s + \gamma_s) \dots\dots\dots (2.4)$$

$\gamma_s$  adalah bias dari unit tersembunyi ke- $s$ .  $b_{net_s}$  adalah *input* dari unit tersembunyi ke- $s$  yang dapat dihitung menggunakan persamaan 2.5.

$$b_{net_s} = \sum_{h=1}^d a_h V_{hs} \dots\dots\dots (2.5)$$

$a_h$  adalah komponen *input* ke- $h$  dari  $x_i$ .  $V_{hs}$  adalah bobot yang menghubungkan unit masukan ke- $h$  dan unit tersembunyi ke- $s$ .

Langkah selanjutnya adalah mencari nilai *general error*. *General error* dari unit keluaran ke- $j$  didefinisikan menggunakan persamaan 2.6 dan *general error* dari unit tersembunyi ke- $s$  didefinisikan menggunakan persamaan 2.7.

$$d_j = \left\{ \begin{array}{l} \left( \frac{1}{|Y_i||\bar{Y}_i|} \sum_{l \in \bar{Y}_i} \exp(-(c_j - c_l)) \right) (1 + c_j)(1 - c_j) \\ \text{if } j \in Y_i \\ \left( -\frac{1}{|Y_i||\bar{Y}_i|} \sum_{k \in Y_i} \exp(-(c_k - c_j)) \right) (1 + c_j)(1 - c_j) \\ \text{if } j \in \bar{Y}_i \end{array} \right\} (2.6)$$

$$e_s = \left( \sum_{j=1}^q d_j W_{sj} \right) (1 + b_s)(1 - b_s) \dots\dots\dots (2.7)$$

Perubahan bobot dihitung menggunakan persamaan 2.8 hingga 2.11.

$$\Delta W_{sj} = \alpha d_j b_s \dots\dots\dots (2.8)$$

$$\Delta \theta_j = \alpha d_j \dots\dots\dots (2.9)$$

$$\Delta V_{hs} = \alpha e_s a_h \dots\dots\dots (2.10)$$

$$\Delta \gamma_s = \alpha e_s \dots\dots\dots (2.11)$$

Apabila telah dilakukan tahap pelatihan untuk setiap pasang data latih maka disebut jaringan telah melakukan 1 epoch tahap pelatihan. Untuk setiap epoch dihitung nilai *global error* menggunakan persamaan 2.12 dan 2.13.

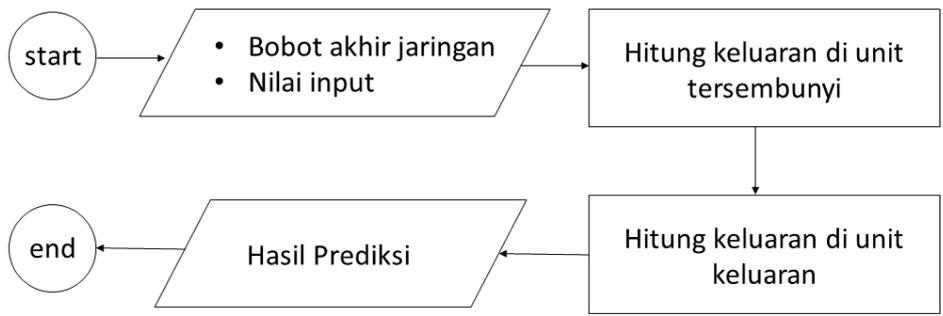
$$E = \sum_{i=1}^m E_i \dots\dots\dots (2.12)$$

$$E_i = \frac{1}{|Y_i||\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)} \dots\dots\dots (2.13)$$

*Global error* adalah jumlah error dari 1 epoch proses pelatihan, sedangkan *training error* adalah nilai rata-rata error dari 1 epoch proses pelatihan. Proses pelatihan berakhir apabila persamaan 2.14 terpenuhi atau epoch sudah mencapai epoch maksimum.

$$E^t - E^{t+1} \leq \epsilon \cdot E^t \dots\dots\dots (2.14)$$

Algoritma pengujian atau algoritma yang digunakan untuk memprediksi label dari data baru hampir sama dengan algoritma pelatihan namun hanya dilakukan tahap-tahap pada fase propagasi maju. Alur algoritma pengujian dapat dilihat pada Gambar 2.3.



Gambar 2.3 Diagram Alur Algoritma Pengujian BP-MLL