**PENERJEMAH BAHASA ISYARAT MENGGUNAKAN METODE *DEEP LEARNING MODEL MULTICHANNEL 2D CONVOLUTIONAL NEURAL NETWORK (M2D CNN)***

**SKRIPSI**



oleh

**Rizkika Zakka Palindungan**

**NIM E41170164**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI JEMBER**

**2020**

# BAB 1. PENDAHULUAN

## Latar Belakang

Manusia adalah makhluk sosial yang saling berkomunikasi dan berinteraksi satu dengan yang lainnya melalui bahasa. Bentuk komunikasi ini biasanya dilakukan secara verbal/lisan, yang artinya komunikasi dengan menggunakan kata-kata. Akan tetapi The World Health Organization (WHO) memberikan pernyataan bahwa lebih dari 5% populasi dunia atau 466 juta orang mengalami gangguan pendengaran (432 juta orang dewasa dan 34 juta anak). Diperkirakan pada tahun 2050 lebih dari 900 juta orang atau satu dari setiap sepuluh orang akan mengalami gangguan pendengaran. Dengan informasi tersebut dapat disimpulkan bahwa akan terjadi peningkatan populasi yang mengalami gangguan pendengaran maupun juga yang mengalami gangguan berbicara, karena orang dengan gangguan pendengaran biasanya akan memiliki gangguan pada kemampuan berbicara. Komunitas yang mengalami gangguan pendengaran disebut sebagai tunarungu sedangkan komunitas yang mengalami gangguan bicara disebut sebagai tunawicara. Tunarungu dan tunawicara memiliki cara berkomunikasi sendiri tanpa menggunakan bahasa lisan yaitu dengan bahasa isyarat. Sedangkan bahasa isyarat ini sulit dipahami oleh masyarakat pada umumnya, dan membuat komunitas tersebut merasa terasingkan di lingkungan sekitarnya. Bahasa isyarat itu sendiri merupakan bahasa yang digunakan untuk berkomunikasi dengan menggunakan gerak bibir dan bahasa tubuh, termasuk ekspresi wajah dan pandangan mata. Selain itu, bahasa isyarat adalah gerakan-gerakan yang sudah disepakati maknanya dan digunakan untuk bertukar informasi.

Untuk mengatasi keterbatasan dalam hal berkomunikasi, teknologi adalah jawabannya. Dengan berkembangnya teknologi visi komputer (*computer vision*), komunikasi antara manusia bisa dijembatani oleh mesin. Teknologi visi komputer (*computer vision*) merupakan salah satu bidang pengetahuan dalam memanfaatkan kecerdasan buatan yang melatih komputer untuk menafsirkan dan memahami dunia visual. Teknologi ini menggunakan gambar digital dari kamera dan bisa juga dari sensor yang kemudian diolah menggunakan model *deep learning*. Maka dari itu mesin dapat secara akurat mengidentifikasi dan mengklasifikasikan objek kemudian bereaksi terhadap apa yang mesin “lihat”. Diharapkan teknologi visi komputer ini bisa dikembangkan untuk membuat sebuah piranti yang mampu menerjemahkan bahasa isyarat ke dalam tulisan bahkan suara. Meskipun harus melewati penelitian dan percobaan yang cukup panjang tetapi teknologi ini pasti akan membantu kehidupan manusia menjadi lebih baik dimasa depan.

Penelitian ini dilakukan dengan pendekatan berbasis visi computer. Pada pendekatan berbasis visi computer digunakan data yang berasal dari tangkapan kamera yang akan menangkap gerakan bahasa isyarat. Sebelum dilakukan proses pengenalan harus melakukan praproses terlebih dahulu. Ada beberapa praproses yang harus dilakukan seperti mengumpulkan data mentah dengan menggunakan kamera, deteksi objek, segmentasi dan konversi warna, thresholding, filtering, operasi morfologi. Setelah data berhasil didapatkan selanjutnya melakukan normalisasi untuk pembuatan model dataset. Normalisasi ini memastikan bahwa data yang berhasil dimodelkan bisa dengan efektif dan efisien digunakan oleh sistem cerdas. Semua praproses tersebut dilakukan untuk memisahkan objek yang sangat penting dari noise. Sampai pada akhirnya sistem komputer yang sudah ditanami dengan teknologi kecerdasan buatan (artificial inteligent) mampu memanfaatkan model dari dataset tersebut dengan maksimal.

Untuk membuat kecerdasan buatan. Salah satu pendekatan yang bisa digunakan adalah menggunakan Jaringan Syaraf Tiruan (JST) yang terinspirasi dari jaringan syaraf pada manusia. Konsep tersebut kemudian dikembangkan lebih lanjut dalam Deep Learning. Deep Learning telah menjadi salah satu topik hangat dalam dunia Machine Learning karena kapabilitasnya yang signifikan dalam memodelkan berbagai data kompleks seperti citra atau visual dan suara. Metode Deep Learning yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah Convolutional Neural Network (CNN). Hal tersebut dikarenakan CNN berusaha meniru sistem pengenalan citra pada visual cortex manusia sehingga memiliki kemampuan mengolah informasi citra.

Sebernarnya konsep penelitian dan pengembangan sistem komputer vision untuk penerjemah bahasa isyarat sudah ada beberapa yang telah dilakukan antara lain : Sistem Pengenalan Bahasa Isyarat Indonesia Dengan Menggunakan Metode Fuzzy K-Nearest Neighbor dengan tingkat akurasi 88% (Gafar, 2017), Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network dengan menganalisis tingkat akurasi dari parameter jumlah total epoch pada model dataset yaitu 25 epoch, 50 epoch, 100 epoch yang memiliki akurasi 67.66%, 89.44%, 96.44% (Bakti., 2019), Model Penerjemah Bahasa Isyarat Indonesia (Bisindo) Menggunakan Convolutional Neural Network dengan melakukan perbandingan 2 buah model dataset yang memiliki akurasi 94.38% dan 30% (Fadillah, 2020). Sebagai pengembangan sistem lebih lanjut, penelitian ini akan menggunakan Bahasa Isyarat Indonesia (BISINDO) sebagai jenis bahasa isyarat yang dipakai didalam dataset. Selanjutnya akan menambah jumlah kamera input menjadi 2 buah untuk mendeteksi tangan dari 2 sisi yaitu sisi depan dan sisi samping hal ini bertujuan untuk meningkatkan kemampuan dan akurasi dalam mendeteksi bahasa isyarat yang lebih kompleks. Pada proses pengolahan data citra dari kamera akan dilakukan penghilangan bagian wajah dan leher dengan menggunakan bantuan Haar Cascade Classifier, kemudian melakukan skin mask atau deteksi warna kulit untuk memfokuskan pada objek tangan. Algoritma yang akan digunakan pada penelitian ini adalah salah satu model pengembangan dari CNN yaitu model arsitektur Multichannel 2D Convolutional Neural Network, algoritma ini dipilih berdasarkan tingkat kemampuan dan akurasi untuk mengolah 2 atau lebih data input citra dari sebuah objek dengan sudut pandang yang berbeda hal ini dibuktikan dari penelitian berjudul A Multichannel 2D Convolutional Neural Network Model for Task-Evoked fMRI Data Classification dengan objek input berupa gambar otak dari 3 sisi yaitu sumbu x, y, z yang memiliki akurasi 83.20% (Hu., 2019).

## Rumusan Masalah

Berdasarkan uraian dari latar belakang, terdapat beberapa permasalahan yang bisa dirumuskan yaitu sebagai berikut :

1. Bagaimana proses ekstraksi fitur warna untuk mendeteksi warna kulit dan dan objek tangan pada data citra dari kamera video?
2. Bagaimana pengolahan dataset yang digunakan untuk data training metode Deep Learning Model Multichannel 2D Convolutional Neural Network?
3. Bagaimana implementasi metode Deep Learning Model Multichannel 2D Convolutional Neural Network dalam menerjemahkan gerakan bahasa isyarat?

## Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengetahui teknik pengolahan citra yang tepat dalam ekstraksi fitur warna dan deteksi objek tangan data citra yang berasal dari kamera.
2. Untuk mendapatkan model dataset dari proses data training yang bisa digunakan oleh mechine learning (M2D CNN) secara efektif dan efisien.
3. Untuk mengetahui hasil metode Deep Learning Model Multichannel 2D Convolutional Neural Network dalam memahami dan menerjemahkan gerakan bahasa isyarat.

## Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut :

1. Meningkatkan efisiensi komputasi pada sistem sehingga memiliki tingkat fungsional dan akurasi yang baik didalam menerjemahkan bahasa isyarat.
2. Memberikan kemampuarn sebuah device untuk belajar mengingat dan memahami gerakan bahasa isyarat manusia.
3. Mampu mengembangkan software untuk memudahkan masyarakat luas dalam memahami bahasa isyarat dengan bantuan mechine learning.

# BAB 2. TINJAUAN PUSTAKA

## State Of The Art

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Penelitian 1 | Penelitian 2 | Penelitian 3 | Penelitian 4 |
| Judul | Sistem Pengenalan Bahasa Isyarat Indonesia dengan Menggunakan Metode Fuzzy K-Nearest Neighbor | Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network | Model Penerjemah Bahasa Isyarat Indonesia (BISINDO) Menggunakan Convolutional Neural Network | Sistem Penerjemah Bahasa Isyarat Menggunakan Metode Deep Learning Model Multichannel 2D Convolutional Neural Network (M2D CNN) |
| Penulis | Agum Agidtama Gafar | Mochamad Bagus Setiyo Bakti dan Yuliana Melita Pranoto | Riestiya Zain Fadillah | Rizkika Zakka Palindungan |
| Tahun | 2017 | 2019 | 2020 | 2020 |
| Metode | Fuzzy K-Nearest Neighbor (FKNN) | Convolutional Neural Network (CNN) | Convolutional Neural Network (CNN) | Multichannel 2D Convolutional Neural Network (M2D CNN) |
| Objek | Citra Bahasa Isyarat Tangan (SIBI) | Citra Bahasa Isyarat Tangan (SIBI) | Citra Bahasa Isyarat Tangan (BISINDO) | Citra Bahasa Isyarat Tangan (BISINDO) |

## Bahasa Isyarat

Bahasa isyarat merupakan jenis komunikasi non verbal karena merupakan bahasa yang tidak menggunakan suara tetapi menggunakan bentuk dan arah tangan, pergerakan tangan, bibir, badan serta ekspresi wajah untuk menyampaikan maksud dan pikiran dari seorang penutur. Kaum tunarungu adalah kelompok utama yang menggunakan bahasa ini. Bahasa isyarat biasanya pengkombinasian dari bentuk, orientasi dan gerak tangan, lengan, tubuh serta ekspresi wajah untuk mengungkapkan isi pikiran (Zuhir dan Amri, 2019).

Di Indonesia, ada dua sistem dari bahasa isyarat yang digunakan yaitu: Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). SIBI adalah bahasa isyarat yang diperkenalkan secara awal oleh Alm. Anton Widyatmoko yang merupakan mantan kepala sekolah SLB/B Widya Bakti di Semarang dalam proses penciptaannya tidak melalui musyawarah dan persetujuan dari Gerakan Kesejahteraan Tunarungu Indonesia (GERKATIN) tetapi berkolaborasi dengan mantan kepala sekolah SLB/B di Jakarta dan Surabaya dengan hasil akhir sebuah kamus SIBI. BISINDO adalah bahasa isyarat yang mengadopsi nilai budaya asli Indonesia dan mudah dapat digunakan untuk berkomunikasi diantara kaum tunarungu dalam kehidupan sehari - hari. Kecepatan dan kepraktisannya dari BISINDO membuat lebih mudah untuk memahami dan mengerti bagi kaum tunarungu walaupun tidak mengikuti faedah tata bahasa dari bahasa Indonesia (Yunanda, 2018).

### Sejarah Bahasa Isyarat Indonesia (BISINDO)

Menurut Laura Lesmana Wijaya selaku Ketua Pusat Bahasa Isyarat Indonesia (Pusbisindo), Bisindo dapat diartikan sebagai sebuah terminologi yang digunakan untuk menunjuk pada bahasa isyarat alami yang digunakan oleh komunitas Tuli di Indonesia. Berdasarkan penjelasan tersebut dapat disimpulkan bahwa sejarah Bisindo sejalan dengan kemunculan bahasa isyarat alami yang terdapat di Indonesia (Yohans, Arjawa dan Punia., 2019).

Kemunculan bahasa isyarat alami diyakini telah berlangsung sejak tahun 1933 ketika sekolah khusus Tuli pertama yaitu Sekolah Luar Biasa (SLB)/B Cicendo, Bandung, Jawa Barat berdiri. Selain itu, terdapat pula sekolah khusus Tuli lainnya yang berdiri pada tahun-tahun berikutnya seperti SLB/B Dena Upakara, Wonosobo, Jawa Tengah (sekolah khusus perempuan) pada tahun 1938, SLB/B Don Bosco, Wonosobo, Jawa tengah (sekolah khusus laki- laki) pada tahun 1955, dan SLB/B Santi Rama (Jakarta) pada tahun 1970 (Tim Produksi Bahasa Isyarat Jakarta, 2014: vii). Penjelasan ini diperkuat dengan keberadaan bahasa isyarat Jakarta yang variasinya berasal dari pencampuran bahasa isyarat asli, termasuk bahasa isyarat yang digunakan oleh orang orang Tuli yang pernah mendapatkan pendidikan formal di sekolah khusus Tuli tersebut (Tim Produksi Bahasa Isyarat Jakarta, 2014: vii).

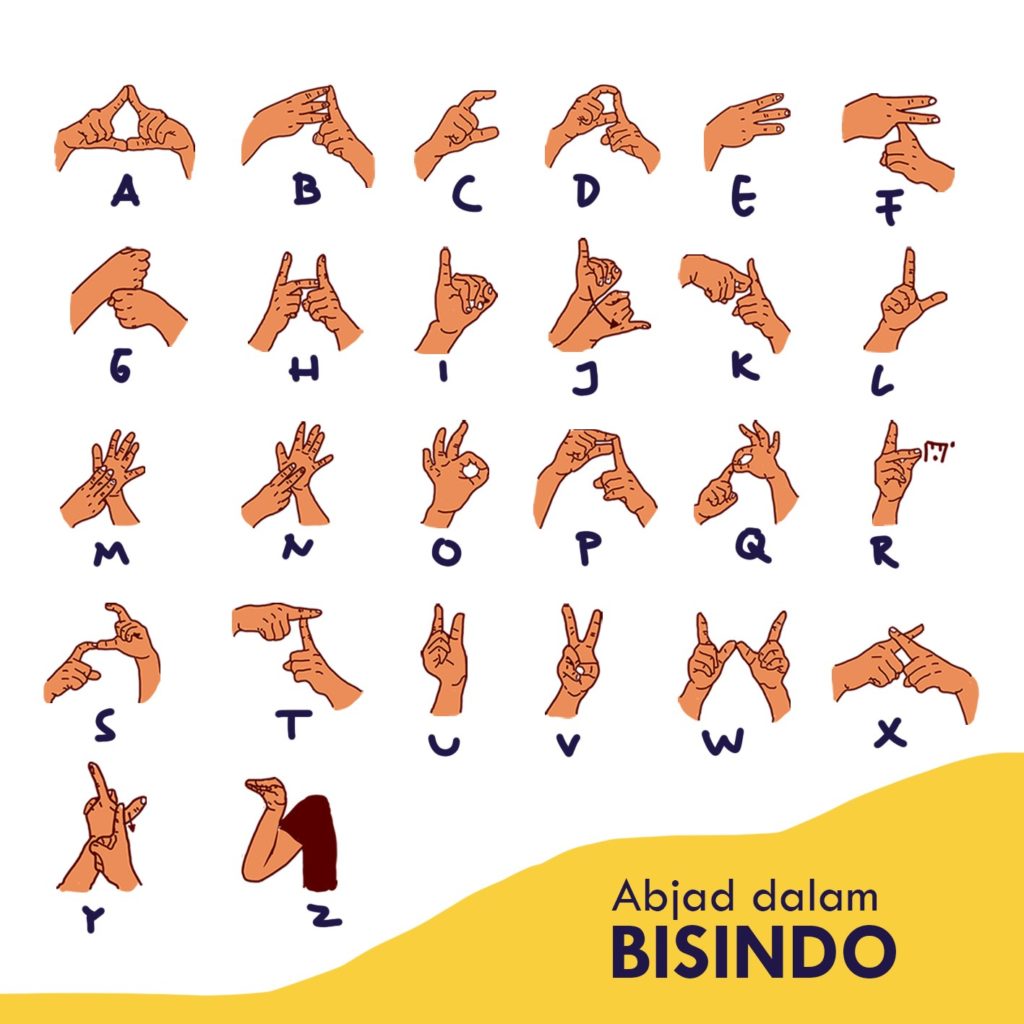
Perkembangan bahasa isyarat alami di Indonesia tidak serta merta mendapatkan pengakuan oleh pemerintah Indonesia. Sistem Isyarat Bahasa Indonesia (SIBI) merupakan sarana komunikasi yang terlebih dahulu diakui oleh pemerintah Indonesia. Pengakuan dan pembakuan atas penggunaan SIBI secara resmi ditetapkan pada tahun 1994 melalui Keputusan Mendikbud RI Nomor 0161/U/1994.

Disebarluaskan dan dibakukannya penggunaan SIBI sebagai sistem isyarat buatan yang bersifat nasional tidak sepenuhnya diterima oleh komunitas Tuli. Komunitas Tuli menilai bahwa keberadaan SIBI tidak merepresentasikan bahasa isyarat asli Indonesia, terdapat berbagai bentuk isyarat yang tidak sesuai dengan isyarat yang berkembang di komunitas Tuli. Salah satu isyarat yang banyak diterapkan pada kamus SIBI, yaitu isyarat yang terdapat pada sistem isyarat American Sign Language (ASL).

Permasalahan dan pertentangan atas penggunaan SIBI menjadi salah satu faktor yang memengaruhi munculnya penggunaan istilah Bisindo. Laura Lesmana Wijaya menjelaskan bahwa penggunaan istilah Bahasa Isyarat Indonesia atau disingkat Bisindo dimulai pada awal tahun 2000. Istilah tersebut muncul melalui pelaksanaan kongres yang dilaksanakan oleh Gerkatin. Penentuan istilah tersebut digunakan untuk menunjuk pada bahasa isyarat alami yang berkembang di komunitas Tuli.

### Isyarat Abjad Dalam Bisindo

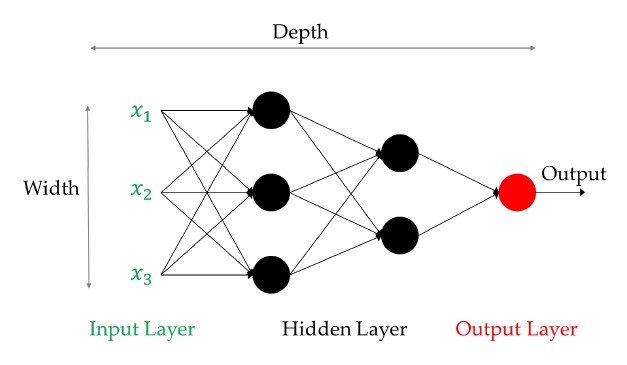
Berikut adalah gambar gerakan bahasa isyarat dari abjad mulai A hingga Z yang diterjemahkan kedalam Bahasa Isyarat Indonesian (BISINDO).



## Artificial Neural Network

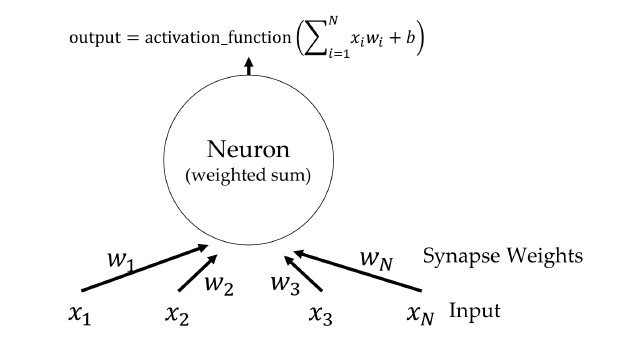
*Artificial neural network* adalah salah satu algoritma *supervised learning* yang populer dan bisa juga digunakan untuk *semi-supervised* atau *unsupervised* *learning* (Amir Atiya, 1994). Walaupun tujuan awalnya adalah untuk mensimulasikan jaringan saraf biologis, jaringan tiruan ini sebenenarnya simulasi yang terlalu disederhanakan, artinya simulasi yang dilakukan tidak mampu menggambarkan kompleksitas jaringan biologis manusia.

*Artificial Neural Network* (selanjutnya disingkat ANN), menghasilkan model yang sulit dibaca dan dimengerti oleh manusia karena memiliki banyak layer (kecuali *single perceptron*) dan sifat **non-linear** (merujuk pada fungsi aktivasi). Pada bidang riset ini, ANN disebut agnostik - kita percaya, tetapi sulit membuktikan kenapa konfigurasi parameter yang dihasilkan *training* bisa benar. Konsep matematis ANN itu sendiri cukup solid, tetapi interpretability model rendah menyebabkan kita tidak dapat menganalisa proses inferensi yang terjadi pada model ANN. Secara matematis, ANN ibarat sebuah graf. ANN memiliki neuron/*node* *(vertex)*, dan sinapsis *(edge)*. Karena memiliki struktur seperti graf, operasi pada ANN mudah dijelaskan dalam notasi aljabar linear. Sebagai gambaran, ANN berbentuk seperti Gambar XX.X (*deep neural network*, salah satu varian arsitektur). *Depth* (kedalaman) ANN mengacu pada jumlah layer. Sementara width (lebar) ANN mengacu pada jumlah unit pada *layer*.



### Single Perceptron

Bentuk terkecil (minimal) sebuah ANN adalah *single perceptron* yang hanya terdiri dari sebuah neuron. Sebuah neuron diilustrasikan pada Gambar XX.X. Secara matematis, terdapat *feature vector* x yang menjadi *input* bagi neuron tersebut (*feature vector* merepresentasikan suatu *data point*, *event* atau *instans*). Neuron akan memproses *input* x melalui perhitungan jumlah perkalian antara nilai *input* dan *synapse weight*, yang dilewatkan pada fungsi non linear. Pada *training*, yang dioptimasi adalah nilai *synapse weight (learning parameter)*. Selain itu, terdapat juga bias *b* sebagai kontrol tambahan (materi *steepest gradient descent*). Output dari neuron adalah hasil fungsi aktivasi dari perhitungan jumlah perkalian antara nilai *input* dan *synapse weight*. Ada beberapa macam fungsi aktivasi, misal ***step function, sign function, rectifier*** dan ***sigmoid function***. Bila diplot menjadi grafik, fungsi ini memberikan bentuk seperti huruf S.



Perhatikan kembali, Gambar XX.X sesungguhnya adalah operasi aljabar linear. Single perceptron dapat dituliskan kembali sebagai XX.X.

Dimana o adalah *output* dan f adalah fungsi non-linear yang dapat diturunkan secara matematis (*differentiable non-linear function* - selanjutnya disebut “fungsi non linear" saja.). Bentuk ini tidak lain dan tidak bukan adalah persamaan model linear yang ditransformasi dengan fungsi non-linear. Secara filosofis, ANN bekerja mirip dengan model linear, yaitu mencari *decision boudary*.

Untuk melakukan pembelajaran *single perceptron, training* dilakukan menggunakan ***perceptron training rule***. Prosesnya sebagai berikut:

1. Inisiasi nilai *synapse weights*, bisa random ataupun dengan aturan tertentu.
2. Lewatkan input pada neuron, kemudian kita akan mendapatkan nilai *output*. Kegiatan ini disebut ***feedforward***.
3. Nilai *output (actual output)* tersebut dibandingkan dengan *desired output*.
4. Apabila nilai output sesuai dengan *desired output*, tidak perlu mengubah apa-apa.
5. Apabila nilai *output* tidak sesuai dengan *desired output*, hitung nilai *error (loss)* kemudian lakukan perubahan terhadap *learning parameter (synapse weight)*.
6. Ulangi langkah-langkah ini sampai tidak ada perubahan nilai *error*, nilai error kurang dari sama dengan suatu threshold (biasanya mendekati 0), atau sudah mengulangi proses latihan sebanyak **T** kali (*threshold*).

Error function diberikan pada persamaan (dapat diganti dengan absolute value) dan perubahan *synapse weight* diberikan pada persamaan , dimana y melambangkan d*esired output*3, melambangkan actual output untuk x sebagai input. disebut sebagai learning rate.

Hasil akhir pembelajaran adalah konfigurasi *synapse weight*. Saat klasifikasi, kita melewatkan *input* baru pada jaringan yang telah dibangun, kemudian tinggal mengambil hasilnya. Pada contoh kali ini, seolah-olah *single perceptron* hanya dapat digunakan untuk melakukan *binary classification* (hanya ada dua kelas, nilai 0 dan 1). Untuk multi-class classification, kita dapat menerapkan berbagai strategi, misal thresholding, i.e., nilai output 0 - 0.2 mengacu pada kelas pertama, 0.2 - 0.4 untuk kelas kedua, dst.

### Multilayer Perceptron

Perhatikan ilustrasi pada Gambar 11.5, multilayer perceptron secara literal memiliki beberapa layers. Pada lecture note ini, secara umum ada tiga layers: input, hidden, dan output layer. Input layer menerima input (tanpa melakukan operasi apapun), kemudian nilai input (tanpa dilewatkan ke fungsi aktivasi) diberikan ke hidden units. Pada hidden units, input diproses dan dilakukan perhitungan hasil fungsi aktivasi untuk tiap-tiap neuron, lalu hasilnya diberikan ke layer berikutnya. Output dari input layer akan diterima sebagai input bagi hidden layer. Begitupula seterusnya hidden layer akan mengirimkan hasilnya untuk output layer.

Kegiatan ini dinamakan *feed forward*. Hal serupa berlaku untuk *artificial neural network* dengan lebih dari tiga *layers*. Parameter neuron dapat dioptimisasi menggunakan metode *gradient-based optimization* Perlu diperhatikan, MLP adalah gabungan dari banyak fungsi *non-linear*. Gabungan banyak fungsi *non-linear* ini lebih hebat dibanding single perceptron. Masing-masing neuron terkoneksi dengan semua neuron pada *layer* berikutnya. Konfigurasi ini disebut sebagai fully connected. MLP pada umumnya menggunakan konfigurasi *fully connected*.