

BAB II

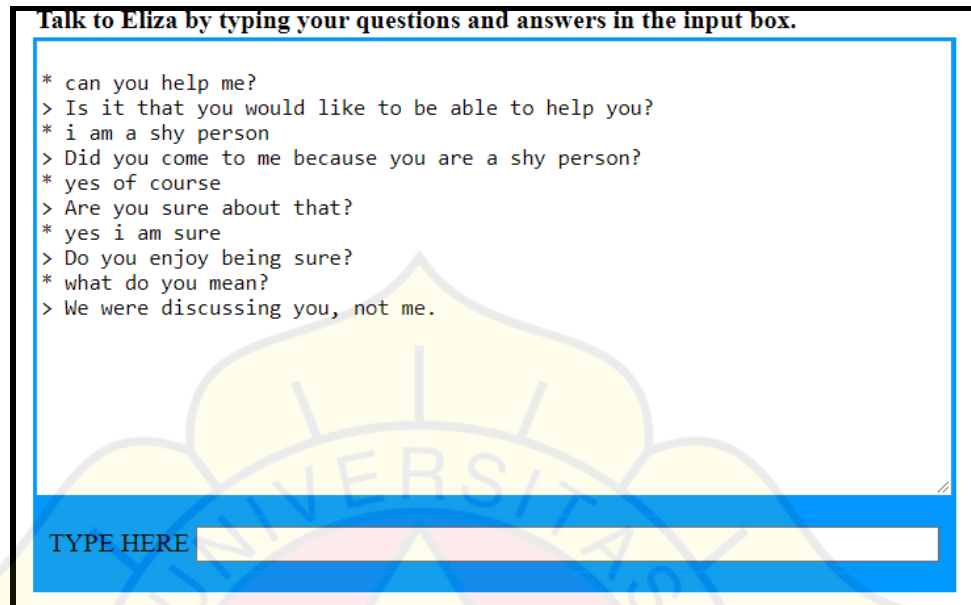
LANDASAN TEORI

2.1 Chatbot

Chatbot merupakan program manusia yang mensimulasikan dialog cerdas antara satu atau lebih pengguna menggunakan teks. *Chatbot* dapat digunakan pada percakapan yang berskala kecil atau dapat juga pada percakapan yang berskala sedang tergantung pada pertanyaannya. *Chatbot* adalah salah contoh dari *Artificial Intelligence* (AI).

Chatbot adalah sebuah kecerdasan buatan dimana perangkat lunak dapat mensimulasikan percakapan dengan pengguna dalam bahasa natural, yang dimaksudkan dalam bahasa natural adalah bahasa yang digunakan oleh manusia dalam komunikasi sehari-hari seperti bahasa Indonesia, Inggris, Jepang, dan lain-lain. Salah satu *chatbot* yang terkenal adalah Eliza. Eliza adalah program *Natural Language Processing* yang dikembangkan pada tahun 1964 – 1966 di MIT Artificial Intelligence Laboratory oleh Joseph Weizenbaum, program ini menggunakan metodologi pencocokan pola dimana setiap kata yang ketik oleh pengguna memiliki bobot tertentu yang nantinya di proses oleh Eliza. Banyak variasi dari kode Eliza yang asli dibuat dan yang paling banyak dikenal adalah DOCTOR, *chatbot* ini dibuat untuk merespon seperti seorang psikoterapis dimana *chatbot* ini akan melemparkan pertanyaan kembali kepada pengguna. Eliza adalah salah satu *chatbot* pertamadiembangkan, *chatbot* ini juga menjadi tes awal untuk *Turing Test*, di *Turing Test* ini mesin akan dicoba untuk menunjukkan perilaku

cerdas yang setara, atau tidak ada bedanya dengan perilaku manusia, tetapi Eliza gagal jika diberikan pertanyaan- pertanyaan yang rumit.



Gambar 2.1 Program ELIZA (John McDaid,2022)

Pelayanan yang diberikan oleh kampus dengan media *whatsapp* tidak berjalan 24 jam. Hal ini membuat pelayanan kurang efektif karena terdapat keterbatasan jam kerja yang membuat pertanyaan-pertanyaan diluar jam kerja tidak mendapatkan respon. Untuk itu dibutuhkan *chatbot* yang menggantikan peran dari *customer service*. Penelitian oleh Hormansyah & Utama (2018), mereka mengimplementasikan *chatbot* menggunakan metode *TF-IDF* dan *Cosine Similarity* pada *website* layanan kesehatan Kota Malang. Pengimplementasian *chatbot* karena dapat memudahkan dalam mencari data karena saat ini data harus dicari satu persatu. Dengan adanya *chatbot* pengunjung dapat melakukan tanya jawab dan mendapatkan informasi tentang pelayanan kesehatan di Kota Malang.

Perkembangan pariwisata semakin pesat setiap tahunnya seiring dengan banyaknya tempat wisata baru yang dibuka. Salah satu kota yang terkenal dengan perkembangan tempat wisata adalah kota Yogyakarta. Yogyakarta memiliki tempat wisata yang beragam dan membutuhkan informasi yang dapat diakses oleh wisatawan lokal dan mancanegara. Salisah et al., (2020), membuat *chatbot* untuk memudahkan wisatawan mencari informasi wisata yang ada di Yogyakarta. Algoritma yang digunakan untuk membangun chatbot menggunakan *Boyer-moore*. Algoritma *Boyer-moore* bekerja dengan cara membandingkan pola atau kata dari kanan ke kiri, sehingga algoritma tersebut menjadi algoritma pencarian yang baik. Hasil pengujian pada chatbot dapat dikatakan bekerja dengan baik.

2.2 Natural Language Processing

Natural Language Processing atau bisa disingkat dengan NLP adalah bidang ilmu percampuran antara ilmu komputer, kecerdasan buatan, dan linguistik. NLP mengangkut kepada interaksi manusia dan komputer, dimana komputer akan memproses dan menganalisa data *natural language* dalam jumlah besar. Hasil yang diharapkan dari proses NLP adalah komputer dapat paham isi dari dokumen yang diterimanya, seperti mengerti konteks dari kata-kata yang ada ataupun hubungan antara kata-kata yang terdapat di dalam dokumen. Sejak lahirnya NLP sekitar tahun 1950-an, NLP semakin banyak digunakan dalam berbagai bidang seperti ritel, perawatan kesehatan, keuangan, hukum, pemasaran, manusia sumber daya, dan banyak lagi.

Natural Language Processing (NLP) adalah bidang kecerdasan buatan yang digunakan untuk menganalisis dan meniru bahasa manusia secara otomatis (Cambria & White, 2014). EYD atau Ejaan Yang Disempurnakan adalah aspek

penting dalam pembuatan teks *essay*, jurnal, dan lainnya. Wangsanegara & Subaeki (2015) mereka melakukan penelitian dengan mengidentifikasi kesalahan dalam penulisan huruf kapital atau tanda baca dengan menggunakan implementasi *Natural Language Processing* (NLP) dan metode *Fuzzy Logic Tsukamoto*. Jumlah kata yang digunakan sebanyak 31.759 kata dan dibangun menggunakan MySQL dan PHP. Hasil dari pengujian dilakukan terhadap 20 abstrak skripsi dengan hasil kesesuaian identifikasi sebesar 70%.

Kota Bandung merupakan kota wisata dengan banyak tempat wisata yang ada. Informasi tentang tempat wisata pada kota Bandung dapat diakses pada *website* pariwisata kota Bandung. S. C. P & Afrianto (2015) melakukan wawancara kepada pengunjung *website* pariwisata kota Bandung dan mendapatkan hasil wawancara pengunjung harus mencari satu per satu tempat wisata pada *website* pariwisata kota Bandung. Dari permasalahan tersebut, mereka membuat sebuah aplikasi *chatbot* yang interaktif. Aplikasi *chatbot* menggunakan bahasa manusia dengan menggunakan metode *Natural Language Processing* (NLP) dan *Unified Model Language* (UML). Hasil pengujian yang didapat adalah wisatawan setuju bahwa aplikasi *chatbot* dapat membantu wisatawan mencari informasi tempat wisata di Kota Bandung.

Opini publik banyak sekali ditemukan pada media sosial, salah satunya pada twitter. Dalam penelitian yang dilakukan oleh Munasatya & Novianto (2020), mereka menggunakan *Natural Language Processing* (NLP) agar data dapat dipahami oleh komputer dan dimasukkan ke dalam model klasifikasi menggunakan model *Multi-Layer Perceptron* yang berguna untuk mendapatkan hasil prediksi akurasi. Objek penelitian ini menggunakan aplikasi *twitter* untuk

Analisis Sentimen opini publik. Untuk mendapatkan visualisasi data menggunakan *Supervised Learning* dengan hasil pengujian yang didapat untuk nilai akurasi sebesar 93,26%.

Ada beberapa proses dasar dalam mengelola data dalam bentuk teks dalam NLP

2.2.1 Tokenization

Tokenization adalah proses dimana memecah kata, kalimat, ataupun seluruh dokumen ke dalam bentuk yang lebih kecil yang disebut tokens. Proses ini membantu dalam penafsiran teks agar lebih mudah di proses.berikut ini contoh dari *word tokenize*



```
27 | initoken = "Aku mau daftar di kampus universitas darma persada"
28 | print(initoken)
29 | cobatokenini = tokenize(initoken)
30 | print(cobatokenini)
31 |
```

TERMINAL OUTPUT JUPYTER DEBUG CONSOLE PROBLEMS 2

```
(andialvin3) Mys-MacBook-Pro:chatbotvinindo2 mymac$ python nltk_utils.py
Aku mau daftar di kampus universitas darma persada
['Aku', 'mau', 'daftar', 'di', 'kampus', 'universitas', 'darma', 'persada']
(andialvin3) Mys-MacBook-Pro:chatbotvinindo2 mymac$
```

Gambar 2.2 Contoh *Word Tokenization*

2.2.2 Stemming

Stemming adalah proses yang bertujuan untuk menghapus atau menghilangkan imbuhan dan mengubah suatu kata menjadi ke bentuk dasarnya seperti kata “mendaftar” dan kata “didaftar” akan menjadi kata dasarnya “daftar” dengan memotong imbuhan dari kedua kata tersebut. Stemming dibutuhkan agar mengurangi kata yang sama dengan imbuhan yang berbeda yang dapat

mengganggu keakuratan model machine learning atau mengganggu dalam pencarian kata.

```
32 words = ["mendaftar","didaftar","mendaftarkan"]
33 stemmed_words = [stem(w) for w in words]
34 print(stemmed_words)
35
36 sentence = 'Saya Hari ini mengkonfirmasi pembayaran ulang'
37 output = stemmer.stem(sentence)
38
39 print(output)
40
```

TERMINAL OUTPUT JUPYTER DEBUG CONSOLE PROBLEMS 2

```
(andialvin3) Mys-MacBook-Pro:chatbotvinindo2 mymac$ python nltk_util
['daftar', 'daftar', 'daftar']
saya hari ini konfirmasi bayar ulang
(andialvin3) Mys-MacBook-Pro:chatbotvinindo2 mymac$
```

Gambar 2.3 Contoh *Stemmer* Kalimat Bahasa Indonesia

2.2.3 Bag of words

Bag-of-words (BoW) merupakan representasi sederhana yang digunakan pada *natural language processing* (NLP) dan *information retrieval* (IR), juga dikenal sebagai model *vector space* (McTear, Callejas, & Griol, 2016). Dalam model ini, sebuah teks yang berupa kalimat ataupun dokumen diwakili sebagai kantung (*bag*) multiset dari kata-kata yang terkandung di dalamnya, tanpa memandang urutan kata dan tata bahasa namun tetap mempertahankan keberagamannya. Definisi lain untuk BoW adalah sebuah model yang mempelajari sebuah kosakata dari seluruh dokumen, lalu memodelkan tiap dokumen dengan menghitung jumlah kemunculan setiap kata (Deepu, Pethuru, & Rajaraajeswari, 2016)

2.3 Jaringan Syaraf Tiruan

Dalam bidang ilmu pengetahuan jaringan syaraf (neural network) sudah sejak lama dibicarakan banyak orang. Mulai dikenal akhir tahun 1940-an, jaringan

syaraf masuk dalam blok perkembangan teknologi komputer. Meskipun begitu, perkembangan teknologi komputer yang menjadi penghambat perkembangan ilmu jaringan syaraf. Pada saat sekarang ini, neural network dapat menjawab berbagai permasalahan seperti telekomunikasi, industry maupun informasi. Jaringan syaraf adalah sistem pengolahan informasi yang didasari fisolofi struktur perilaku syaraf makhluk hidup. Dalam arsitekturnya, jaringan syaraf mempelajari bagaimana menghasilkan keluaran yang diinginkan pada saat diberikan sekumpulan masukan (Siswanto, 2010).

Salah satu sistem penjurusan yang ada pada program studi adalah peminatan. Peminatan berfungsi untuk lebih fokus memilih satu bidang yang disukai sebagai penentu topik dari skripsi. Dalam penelitian oleh Riani et al., (2021), mereka melakukan penelitian terhadap program studi Sistem Informasi Universitas Telkom tentang penjurusan mahasiswa. Program studi Sistem Informasi menyediakan lima bidang peminatan yang nantinya dapat dipilih oleh mahasiswa semester tujuh. Mahasiswa masih kesulitan dalam memilih bidang peminatan yang dipilih, dan lebih menyukai bidang yang dianggap mudah. Oleh karena itu dilakukan penelitian untuk proses seleksi pemilihan peminatan mahasiswa program studi Sistem Informasi dengan menggunakan *Neural Network*. *Neural Network* digunakan karena mampu untuk mengklasifikasi data yang sangat kompleks. Hasil yang didapat adalah nilai AUC sebesar 0.9950 dan nilai akurasi sebesar 0.9863.

Indonesia tanaman tomat banyak dijumpai dan dibudidayakan oleh masyarakat karena iklim yang mendukung dan harga yang tomat yang stabil. Tetapi, terdapat beberapa kendala bagi para petani dalam budidaya tomat karena

rentan diserang penyakit yang mengakibatkan petani kesulitan dalam mendeteksi penyakit tanaman tomat. Dalam penelitian yang dilakukan oleh Putri (2021), menggunakan metode *Neural Network* dan algoritma *backpropagation* dalam membantu petani untuk mengidentifikasi penyakit pada daun tanaman tomat. Hasil yang didapat untuk akurasi sebesar 78% , presisi sebesar 0,78 dan data klasifikasi bernilai benar sebanyak 117 data.

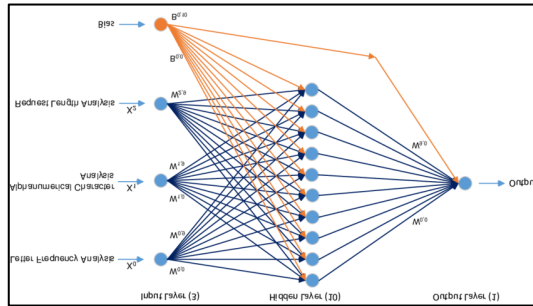
Indonesia adalah negara hukum dengan banyak hukum dan peraturan. Dalam penelitian Nia Agustina Purwitasari & Soleh (2022), Mereka membuat sistem untuk memudahkan masyarakat mendapatkan informasi tentang hukum Indonesia saat ini dengan menerapkan chatbot yang disebut "Peraturanku" untuk menghindari kebingungan tentang hukum dan peraturan. Mereka menggunakan *Neural Network* dengan pendekatan *Natural Language Processing* (NLP). Hasil uji coba dengan data sebanyak 35 percakapan dan berhasil menjawab percakapan dengan benar sebanyak 33 percakapan dengannilai akurasi 94,28% dan kesalahan 5,71%.

Jaringan syaraf tiruan (JST) merupakan suatu sistem pemroses informasi yang memiliki karakteristik yang mirip dengan jaringan syaraf biologi (Fausset, 1994). Simon Haykin menyatakan bahwa jaringan syaraf tiruan adalah sebuah mesin yang dirancang mirip dengan cara kerja otak manusia untuk menyelesaikan suatu permasalahan atau mengerjakan tugas-tugas tertentu. Mesin ini memiliki kemampuan untuk menyimpan pengetahuan berdasarkan pengalaman dan menjadikan pengetahuan itu menjadi bermanfaat (Desiani dan Arhami, 2006).

Jaringan syaraf tiruan (JST atau Artificial Neural Network), adalah suatu sistem komputasi yang mana arsitektur dan operasinya terinspirasi dari pengetahuan tentang sel syaraf biologis di dalam otak manusia, dan merupakan representasi buatan dari otak yang selalu menstimulasi proses pembelajaran pada otak manusia tersebut (Hermawan, 2006).

Jaringan Syaraf Tiruan merupakan cabang dari AI (Artificial intelligence). Jaringan syaraf tiruan merupakan representasi buatan dari otak manusia yang selalu mencoba untuk menstimulasikan proses pembelajaran pada otak manusia. Kata buatan disini digunakan karena jaringan syaraf diimplementasikan dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran. (Kusumadewi, 2003 dikutip oleh Fakhrurriqfi dkk, 2013).

Menurut Hecht-Nielsen pada tahun 2006 menyatakan sistem syaraf tiruan sebagai berikut: neural network merupakan suatu struktur pemrosesan informasi yang terdistribusi dan bekerja secara paralel yang terdiri dari elemen-elemen pemroses yang saling terkoneksi dengan alur sinyal searah. Dan setiap elemen pemroses yang terkoneksi memiliki keluaran tunggal yang bercabang ke sejumlah koneksi kolateral yang diinginkan. Keluaran dari elemen pemroses dapat berupa sembarang jenis persamaan matematis yang diinginkan, keluaran bergantung pada nilai masukan yang diperoleh melalui koneksi dan nilai yang tersimpan dalam memori lokal.

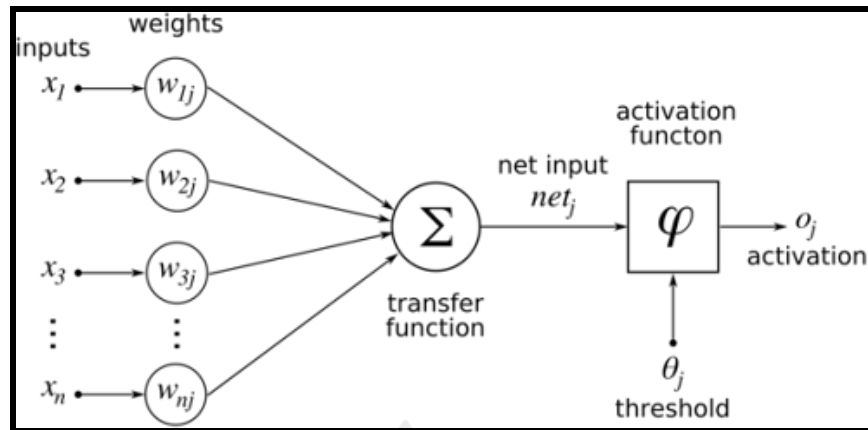


Gambar 2.4 Model Dasar *Neural Network* (Adem Tekerek, 2019)

Gambar memperlihatkan model dasar dari neural network, yang terdiri dari *input layer*, *hidden layer*, dan *output layer*. Gambar tersebut juga memperlihatkan hubungan antara *neuron* dimana dihubungkan dengan semua *neuron* dari *layer* sebelumnya, ini dinamakan *fully connected layer*, *layer* jenis ini umum digunakan dalam *neural network*.

2.3.1 Perceptron

Pada jaringan saraf pada otak manusia terdapat *neuron* yang bekerja sebagai menghantar rangsangan dari reseptor ke otak atau sebaliknya, *neuron* pada *neural network* juga bekerja mirip seperti *neuron* yang terdapat pada otak manusia. *Neuron neural network* dimodelkan oleh McCulloch-Pitts ke dalam sel saraf buatan yang disebut *perceptron*.



Gambar 2.5 Model Matematis Neural Network (Francisco Martínez-Álvarez,2015)

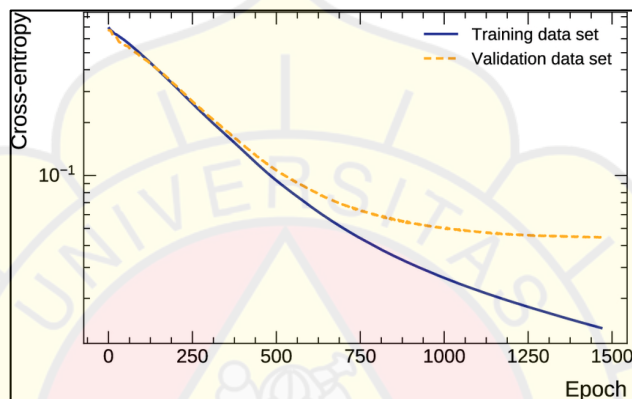
Seperti gambar bobot-bobot sinaptik dinotasikan dengan $\omega_1, \omega_2, \dots, \omega_n$, untuk bias dinotasikan dengan b_n . Sinyal masukan dinotasikan dengan x_1, x_2, \dots, x_n (Suyanto, 2019). Dari gambar tersebut dapat dilihat bahwa *neural network* mempunyai beberapa *input(x)* yang saling terkoneksi dengan masing- masing bobot(ω) tertentu. Bobot akan dikalikan dengan masing-masing *input* yang menghubungkannya. Bias dan bobot adalah parameter yang dapat mempengaruhi *output* dari *perceptron*, dalam *neural network* nantinya akan ada puluhan bahkan sampai ribuan parameter bias dan bobot yang di optimisasi pada saat *training* model.

2.3.2 Fungsi Aktivasi

Fungsi aktivasi digunakan untuk membantu *neural network* mengenali pola data yang rumit, fungsi ini ditambahkan pada *output* dari *neuron* atau *layer neuron* untuk memodifikasi nilai outputnya. Nilai *output* dari hasil fungsi aktivasi akan menjadi *output* dari *neural network* ataupun menjadi input untuk *layer* selanjutnya.

2.3.3 Loss Function

Fungsi *loss* atau *loss function* adalah algoritma untuk melihat *error output* yang dihasilkan oleh *neural network* terhadap hasil yang diharapkan pada saat *training*. Semakin tinggi nilai *loss* yang dihasilkan maka dapat disimpulkan prediksi yang dihasilkan *neural network* kurang baik, sebaliknya maka hasil dapat dikatakan mendekati atau sama dengan hasil yang diharapkan.



Gambar 2.6 Grafik *loss function* (Maurizio Pierini,2019)

Ada beberapa jenis *loss function* yang sering digunakan beberapa diantaranya adalah *mean squared* dan *categorical cross-entropy loss*. *Mean squared error* (MSE) dapat dihitung dengan menghitung kuadrat dari perbedaan dengan prediksi *neural network* dengan hasil yang diharapkan untuk seluruh *output* dari *neuron*, kemudian nilai tersebut dicari rata-ratanya.

Berikut adalah persamaan dari *mean squared error*:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Dengan n adalah jumlah sampel, \hat{Y} adalah vektor nilai perkiraan, dan Y adalah vektor nilai sebenarnya.

Categorical cross-entropy biasanya digunakan untuk menghitung *loss* untuk *output neural network* bersifat distribusi peluang seperti *softmax*. Berikut adalah persamaan dari *categorical cross-entropy*:

$$CCE = \sum y_{i,j} \log(\bar{y}_{i,j})$$

2.4 Metode Feed Forward Neural Network

Feed forward neural network (FFNN) merupakan salah satu jenis utama dari *neural network*, dan metode FFNN merupakan metode yang paling banyak digunakan untuk melakukan prediksi (Meinanda dkk, 2009). FFNN diperkenalkan oleh M. Minsky dan S. Papert pada tahun 1969. Jaringan pada model FFNN menggunakan algoritma *backpropagation* dikembangkan oleh David E. Rumelhart, Geoffrey E. Hinton, dan Ronald J. Williams pada tahun 1986. Pengembangan algoritma ini menjadi awal kebangkitan riset di bidang *neural network* (Fajar dan Jamitko, 2011). Algoritma *backpropagation* meliputi tiga tahap, yaitu umpan maju (*feedforward*) dari pola *input*, penghitungan dan propagasi balik (*backforward*) dari *error*, serta penyesuaian bobot. Pada tahap *feedforward* setiap unit *input* menerima sinyal *input*

(x_i) dan menyebarkannya ke unit tersembunyi z_1, \dots, z_p . Kemudian semua unit tersembunyi menghitung aktivasinya lalu mengirimkan sinyal (z_j) ke unit *output*. Setelah itu, unit *output* menghitung aktivasinya dan memprediksikan nilai variabel respons (Warsito, 2006).

Diantara berbagai model *neural network*, *feed forward neural network* (FFNN) merupakan model yang lebih sering digunakan karena dikenal memiliki kemampuan pendekatan yang baik dan bersifat universal (Handaga dan Asy'ari,

2012). Selain itu, metode FFNN telah dikenal akan keunggulannya yaitu memiliki nilai prediksi yang sangat mendekati nilai aktualnya sehingga menghasilkan galat yang kecil serta memiliki kemampuan untuk mendeteksi atau melakukan analisis untuk permasalahan yang bersifat sangat kompleks. Metode FFNN tidak memiliki syarat atau asumsi tertentu (Setiawan, 2008).

Pada FFNN, *neuron-neuron* diaktifkan dengan menggunakan fungsi aktivasi yang dapat dideferensiasikan seperti sigmoid dengan menggunakan Persamaan (Kusumadewi, 2004) sebagai berikut:

$$Y = f(x) = \frac{1}{1 + e^{-\sigma x}}$$

dengan: $f'(x) = \sigma f(x)[1 - f(x)]$

2.5 Arsitektur Metode Feed Forward Neural Network

Pada FFNN, setiap unit yang ada di *input layer* terhubung dengan unit yang ada di *hidden layer*, dan setiap unit yang ada di *hidden layer* terhubung dengan unit yang ada di *output layer*.

Jaringan syaraf tiruan FFNN terdiri dari banyak lapisan, yaitu :

1. Lapisan masukan (*input layer*)

Lapisan masukan terdiri dari unit-unit *input*, mulai dari unit *input* 1 sampai unit *input* n.

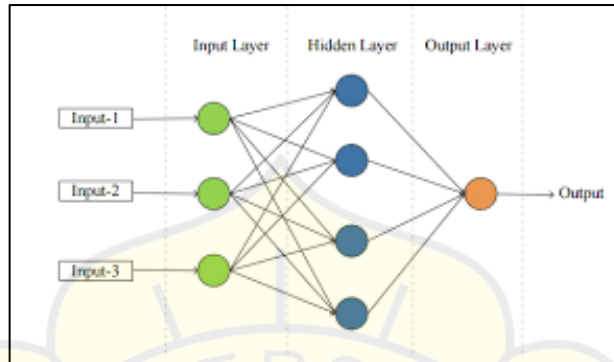
2. Lapisan tersembunyi (*hidden layer*)

Lapisan tersembunyi terdiri dari unit-unit tersembunyi, mulai dari unit tersembunyi 1 sampai unit tersembunyi p.

3. Lapisan keluaran (*output layer*)

Lapisan keluaran terdiri dari unit-unit keluaran, mulai dari unit keluaran 1 sampai unit keluaran m,n,p

Arsitektur model FFNN akan diilustrasikan dengan gambar berikut:



Gambar 2.7 Arsitektur model (Lamyaa Dawood,2017)