

BAB 2
TINJAUAN PUSTAKA
DAN
LANDASAN TEORI

2.1 Tinjauan Pustaka

Data mining atau penambangan data adalah proses ekstraksi dari prediksi informasi yang tersembunyi dari sebuah kumpulan data. Hasil penambangan data dapat berupa pengetahuan baru yang didapatkan dari proses penambangan data tersebut. Dalam penerapannya, penambangan data ini memiliki dua hasil utama yaitu sebuah prediksi dan sebuah deskripsi (informasi atau pengetahuan baru). Prediksi ini adalah sebuah proses pengolahan data yang ada pada dataset yang akan menghasilkan sebuah prediksi nilai di masa yang akan datang. Sedangkan deskripsi lebih fokus kepada pencarian pola - pola dan informasi yang menggambarkan keadaan data agar dapat disajikan kepada pengguna data selanjutnya (Geetha, et al., 2008).

Salah satu proses yang dapat dilakukan pada *data mining* adalah klasifikasi. Metode *Artificial Neuron Network* (ANN) merupakan salah satu metode untuk melakukan klasifikasi. Dalam penelitian tentang Jaringan Syaraf Tiruan Dengan Algoritma Backpropagation untuk Penentuan Kelulusan Sidang Skripsi (Matondang, 2013) diketahui bahwa algoritma pada metode ANN yang sering digunakan adalah algoritma Backpropagation. Pada metode dengan algoritma ini menggunakan metode ini, kunci keberhasilan dari hasil dari ANN terdapat pada pelatihan data. Pada penelitian ini pelatihan data masih dilakukan pada iterasi pertama saja sehingga belum dapat mencapai nilai *error* sesuai dengan target peneliti. Pada penelitian tentang Analisis Metode

Jaringan Syaraf Tiruan Backpropagation untuk Pengenalan Sel Kanker Otak (Handayani, 2013), dalam menggunakan metode ANN, selain pelatihan data, hal penting yang perlu diperhatikan adalah arsitektur ANN yang dibuat dan juga bobot yang akan digunakan untuk membandingkan hasil yang keluar. Hasil dari penelitian kurang maksimal karena bobot dan arsitektur yang digunakan kurang sesuai. Permasalahan mengenai bobot pada metode ANN dapat diatasi pada penelitian tentang Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan *Certainty Factor* (Pakaja, et al., 2012). Pada penelitian ini digunakan metode *Certainty Factor* yang dikombinasikan pada ANN. Fungsi *Certainty Factor* adalah memberikan nilai pembanding yang terbaik yang digunakan untuk melakukan koreksi bobot dari hasil yang dikeluarkan ANN. Dengan metode ini, tingkat akurasi hasil data akan bertambah. Salah satu kegunaan lain dari penggunaan metode ANN yaitu digunakan untuk peramalan data yang akan datang. Pada penelitian tentang Penerapan Algoritma Jaringan Saraf Tiruan Backpropagation dalam Memprediksi Tingkat Suku Bunga Bank (Anwar, 2011), diketahui bahwa parameter pelatihan data untuk memprediksi sesuatu hal dapat berpengaruh pada hasil prediksi. Pada penelitian ini keakuratan prediksi mencapai 94% yang menandakan bahwa akurasi sudah sangat baik.

Dari penelitian yang pernah dilakukan terhadap data EEG dengan melakukan klasifikasi berdasarkan penyakit - penyakit yang dapat dideteksi melalui data EEG diketahui bahwa penyakit - penyakit yang dapat diklasifikasikan adalah penyakit epilepsi, *Mood Disorder*, *Dementia*, *Schizophrenia* dan *Attention*

Deficit Hyperactivity Disorder. Dari penelitian tersebut klasifikasi data EEG menggunakan Weka dan memanfaatkan algoritma J48 yang memberikan hasil akurasi tes untuk mengidentifikasi negatif orang - orang yang tidak memiliki penyakit (spesifisitas) berada diantara 94% hingga 100% sedangkan akurasi tes untuk mengidentifikasi positif orang - orang memiliki penyakit (sensitifitas) berada diantara 70% hingga 100% (Pandey & Kundra, 2014). Angka pasti akurasi dari spesifisitas dan sensitifitas deteksi penyakit epilepsy melalui data EEG ini dapat diketahui pada penelitian *Epileptic EEG Detection Using Neural Networks and Post-Classification* dimana hasil akurasi spesifisitas berada pada angka 99.99% dan untuk akurasi sensitifitas berada pada angka 91.29% (Patnaik & Manyam, 2008) . Pada penelitian tersebut digunakan metode Wavelet Transfor untuk melakukan ekstraksi fitur dan proses klasifikasi menggunakan ANN dengan algoritma *backpropagation*. Pada penelitian tentang *The Use Of Wavelet Power Spectrum For Detection And Identification Of Thinking-Induced Eeg Signals*, data EEG yang diambil dari hasil perekaman alat dengan 20 kanal sinyal dapat diklasifikasikan ke dalam dua kelas yaitu kondisi rileks dan tidak rileks. Proses analisis dan klasifikasi menggunakan spektral daya yang sebelumnya menggunakan transformasi wavelet sebagai pra pengolahan data EEG dimana dalam proses ini dapat memberikan keberhasilan klasifikasi sebesar 85% untuk kondisis rileks dan 83% untuk kondisi non rileks. Pada penelitian ini dilakukan juga klasifikasi tanpa melewati proses pra pengolahan data, dimana hasil klasifikasi tanpa proses pra pengolahan data dapat menurunkan keberhasilan klasifikasi menjadi 58% untuk kondisi rileks dan 60% untuk kondisi tidak rileks. Dari

penelitian ini dapat disimpulkan bahwa proses pra pengolahan data sebelum proses klasifikasi dapat memberikan peningkatan keberhasilan klasifikasi. Hal ini dikarenakan pada metode transformasi wavelet dapat mengatasi sinyal yang non stasioner (Djamal, et al., 2005). Pada penelitian sebelumnya ini tidak dilakukan eksperimen lebih jauh mengenai keterkaitan unsur waktu pada data EEG yang dimiliki dalam melakukan klasifikasi.

Proses penambangan data sangat berpengaruh pada jenis datanya. Pada data EEG, data yang dihasilkan oleh EEG bersifat data temporal sehingga proses penambangan data dilakukan menggunakan teknik *temporal data mining*. Pada teknik ini, pemrosesan data akan dilakukan berdasarkan kurun waktu tertentu. Pada proses klasifikasi epilepsi berdasarkan data EEG, sebuah data tertentu belum dapat menunjukkan apakah data tersebut memiliki gejala penyakit epilepsi atau belum. Butuh beberapa data dalam waktu tertentu untuk dapat menentukan apakah data tersebut adalah data yang memiliki gejala epilepsi atau tidak. Maka dari itu, eksperimen untuk menentukan waktu dalam melakukan klasifikasi penyakit epilepsi dibutuhkan untuk menemukan batas waktu terbaik. Pada penelitian yang pernah dilakukan menggunakan *Time Domain* dan *Frequency Domain Features* pada data EEG. Sebelum dilakukan pemrosesan data, terlebih dahulu data EEG dimasukkan ke dalam pra proses. Pada bagian ini, data EEG dengan durasi 23.6 detik dilakukan pemotongan dimana tiap segmen pemotongan memiliki durasi 1 detik. Pemotongan ini digunakan untuk melakukan pengenalan dan pembelajaran pada ANN untuk mengenali pola dari data-

data yang dipotong (Srinivasan, et al., 2005). Hasil dari klasifikasi menggunakan memberikan akurasi sebesar 99,6%.

Pada penelitian ini, data yang digunakan adalah data EEG dari dari penderita epilepsi yang disediakan oleh Universitas Bonn. Terdapat beberapa penelitian yang telah menggunakan data ini untuk melakukan penelitian terhadap data EEG. Salah satu penelitian di Indonesia dengan menggunakan data ini adalah penelitian yang berjudul *Klasifikasi Sinyal Eeg Menggunakan Koefisien Autoregresif, F-Score, dan Least Squares Support Vector Machine* (Karyawan et al., 2011). Pada penelitian ini data EEG diklasifikasikan menggunakan LS-SVM dengan kombinasi *autoregresif* sebagai fitur ekstraksi dan *F-score* sebagai fitur seleksinya. Pada penelitian ini menitik beratkan pada klasifikasi orang sehat dengan mata terbuka dan orang dengan penderita epilepsi pada kondisi kejang. Dari penelitian ini didapatkan akurasi klasifikasi sebesar 99,64%. Pada penelitian ini, hanya digunakan 2 dataset dari keseluruhan 5 dataset yang ada pada data yang disediakan. Pada penelitian ini data EEG tidak melakukan eksperimen pada unsur waktu yang terdapat pada data EEG.

Penelitian dengan data yang sama juga pernah dilakukan dengan menggunakan ANN sebagai pengklasifikasi. Pada penelitian ini, kelima dataset digunakan untuk klasifikasi. Dimana masing-masing dataset mewakili masing-masing kelas yang ada. Klasifikasi dilakukan pada kombinasi dataset S dan Z, kemudian dataset S, F, dan Z, serta yang terakhir adalah dataset S, F, Z, N, dan O. Penelitian ini

menggunakan metode *Time-Frequency Analysis* dimana proses klasifikasi yang menggunakan ANN memberikan akurasi terbaik dengan memberikan akurasi sebesar 100% untuk kombinasi pertama, 100% untuk kombinasi kedua, dan 89% untuk kombinasi ke 3 (Tzallas et al., 2009).

2.2 Dasar Teori

2.2.1. Data Mining

Data mining adalah sebuah proses melakukan analisis data dari sudut pandang yang berbeda dan menarik kesimpulan dari hasil analisis menjadi berbagai informasi baru yang penting yang dapat dipakai untuk membentuk pengetahuan baru guna memperoleh sebuah keuntungan (Mabrur & Lubis, 2012) sedangkan pengertian lain dari *data mining* adalah proses menemukan pola yang memiliki arti dan pengetahuan yang ada dari data yang berjumlah besar (Han, Kamber, & Pei, 2012). Analisis data yang dilakukan pada *data mining* merupakan analisis pada data dengan jumlah yang sangat besar dan memiliki berbagai jenis dan bentuk. *Data mining* digunakan untuk mencari dan menemukan hubungan antar data dan juga mengidentifikasi dan mempelajari pola yang muncul dari hubungan antar data tersebut. Pola ditemukan ini akan diambil dan diproses untuk menjadi informasi yang baru dan berharga dari kumpulan data tersebut.

Dari pengertian di atas mengenai *data mining* yang telah dijabarkan oleh para ahli, dapat diketahui bahwa *data mining* adalah suatu pemrosesan data yang melibatkan proses pencarian dan analisis pada suatu kumpulan data yang besar yang akan

menghasilkan suatu pola atau karakter data yang dapat berarti dan dapat digunakan untuk membentuk sebuah informasi atau pengetahuan yang berguna. *Data mining* ini merupakan salah satu tahapan dari proses yang ada pada *Knowledge Discovery in Database* (KDD) yang digunakan untuk mengambil intisari dari model atau pola tertentu yang muncul dari analisis data.

Secara umum, *data mining* dapat dikelompokkan menjadi 2 kategori utama, yaitu *predictive* dan *descriptive* (Han, Kamber, & Pei, 2012) :

a. Predictive

mining adalah proses pencarian pola-pola dari kumpulan data menggunakan variabel tertentu yang mungkin akan terjadi di waktu yang akan datang. Proses ini dilakukan dengan cara melakukan analisis kepada kumpulan data secara urut dan membentuk satu atau beberapa kumpulan model yang dapat digunakan untuk memprediksi karakteristik dari kumpulan data baru. Contoh dari penerapan *predictive data mining* adalah proses klasifikasi, analisis regresi dan analisis trend.

b. Descriptive

Descriptive mining adalah proses pencarian karakter atau ciri-ciri serta informasi tertentu yang penting dalam suatu kumpulan data. Tugas *descriptive mining* adalah melakukan penyelidikan atas informasi yang bisa didapatkan dari kumpulan data tertentu. Informasi yang didapatkan, atau cirri yang didaptkan dari

proses ini perlu dilakukan proses validasi dan pemberian penjelasan terhadap temuan baru yang ada pada data. Proses ini disebut *post-processing*.

Data mining memiliki 3 tujuan utama dari proses yang dilakukan yaitu *explanatory*, *confirmatory*, dan *exploratory* (Hoffer, Ramesh & Topi, 2012). Tujuan *data mining* dalam kaitannya dengan *explanatory* adalah untuk memberikan penjelasan pada suatu kondisi tertentu atau pada suatu proses observasi. Sedangkan tujuan *confirmatory* digunakan untuk melakukan konfirmasi atau menjawab hipotesis yang dibuat. Terakhir adalah *exploratory* yang bertugas untuk menganalisis kumpulan data yang akan diproses. Sebagai salah satu proses analisis, *data mining* memiliki kelebihan dalam melakukan proses analisis. Yang pertama jelas bahwa *data mining* dapat menangani pemrosesan terhadap data yang besar dan kompleks seperti memiliki berbagai macam tipe data dan atribut yang dimiliki. Kelebihan yang penting pada *data mining* adalah kemampuannya untuk mempelajari pengalaman atau proses pembelajaran terdahulu. Dengan kemampuan ini, hasil dan kualitas analisis proses yang dilakukan pada waktu selanjutnya akan meningkat.

Pada *data mining* terdapat beberapa tahapan yang dilakukan untuk memproses kumpulan data. Proses pertama yang dilalui adalah *data selection*. Pada pemilihan data ini adalah proses memilih kumpulan data yang akan dianalisis. Lalu yang kedua adalah *preprocessing* yaitu proses pembuangan data yang ganda, data yang inkonsisten, serta penggabungan

data dengan data lain untuk memperkaya informasi guna keperluan analisis. Proses ketiga adalah *transformasi* yaitu proses pengkodean sesuai dengan kriteria analisis pada data yang telah dipilih. Kemudian proses pencarian pola atau informasi yang baru dengan metode atau teknik tertentu. Proses terakhir adalah proses evaluasi yaitu proses pengkajian hasil penemuan informasi baru dengan fakta atau hipotesis yang sebelumnya telah dibuat lalu dipresentasikan dalam bentuk yang sesuai dengan keinginan.

2.2.2. Artificial Neuron Network (ANN)

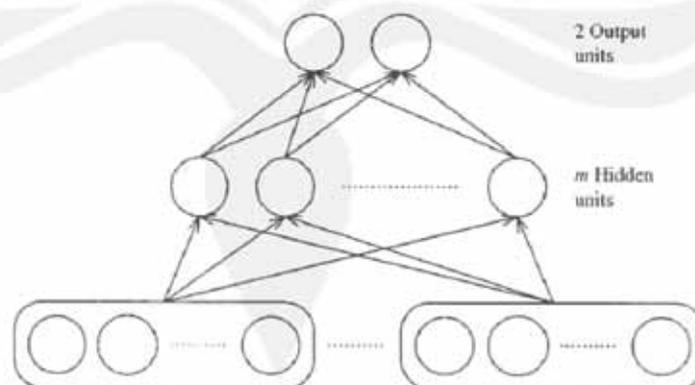
Artificial Neuron Network (ANN) atau Jaringan Saraf Tiruan adalah sebuah konsep rekayasa pengetahuan yang mengadopsi sistem kerja saraf manusia. Metode ini dapat digunakan untuk pengenalan pola, klasifikasi dan peramalan. Dalam desainnya, ANN memiliki 3 bagian yaitu bagian input, bagian pemrosesan dan bagian output (Prasetyo, 2012). Inputan pada ANN ini dapat berupa vector sehingga perhitungan dalam ANN dapat dilakukan untuk masalah yang kompleks dengan mudah.

Dalam prosesnya, metode ANN ini digunakan untuk melakukan peramalan dan pengenalan pola dalam *data mining*. Untuk melakukannya, ANN memerlukan proses pelatihan agar dapat melakukan prediksi kelas dari suatu data uji coba. Dalam proses penambahan data, ANN menggunakan fungsi aktivasi yang digunakan untuk membatasi keluaran dari bagian pemrosesan atau neuron agar sesuai dengan batasan yang diinginkan. Terdapat berbagai algoritma yang dapat digunakan

untuk menggunakan metode ini. Salah satunya adalah algoritma *Backpropagation*.

Algoritma *Backpropagation* adalah salah satu algoritma yang digunakan untuk melakukan pelatihan pada metode ANN. Algoritma ini bersifat nonlinear yang dapat mengatasi berbagai masalah yang rumit. Algoritma ini memiliki dasar matematis yang tinggi dan dilatih menggunakan metode belajar terbimbing dimana hasil atau tujuannya sudah diketahui sebelumnya. Pada algoritma ini, jaringan akan diberikan sepasang pola yang merupakan masukan dan pola yang diinginkan. Ketika pola dimasukkan ke dalam jaringan maka bobot-bobot akan diubah untuk memperkecil perbedaan pola keluaran dengan pola yang diinginkan. Pelatihan ini dilakukan berulang-ulang sehingga memenuhi pola yang diinginkan,

Algoritma ini mendukung jenis ANN yang bersifat *multi layer* atau biasa disebut *Multi Layer Preceptron* (MLP). Pada algoritma ini terdiri dari 3 layer yaitu layer input, layer tersembunyi dan layer output.



Gambar 2.1. Arsitektur Multi Layer Preceptron (Haselstainer & Pfurtscheller, 2000).

Pada gambar di atas dapat dilihat arsitektur dari algoritma ini. Dicontohkan dengan gambar di atas diketahui terdapat layer output lalu satu layer tersembunyi dan satu layer input. Pada layer tersembunyi ini dapat dibuat lebih dari satu layer. Proses yang berjalan pada algoritma ini adalah nilai dikirim melalui layer input ke layer tersembunyi sampai ke layer output. Pada bagian layer tersembunyi memiliki fungsi aktivasi sigmoid. Pelatihan pada algoritma *Backpropagation* memiliki proses seperti berikut. Yang pertama adalah proses Inisialisasi yang digunakan untuk memberikan nilai awal terhadap nilai yang diperlukan seperti bobot. Lalu yang kedua adalah Aktivasi yang akan melakukan perhitungan aktual output pada layer tersembunyi dan layer output. Selanjutnya adalah Pelatihan Bobot yaitu menghitung error pada layer tersembunyi dan layer output dan mengganti bobot dengan bobot yang baru dan terakhir adalah Iterasi yang akan mengulang proses ini hingga mendapat error yang minimal. Pada algoritma ini, proses pelatihan memiliki dua fase utama. Fase yang pertama, inputan pada algoritma ini diberikan pada layer input pada jaringan. Kemudian inputan akan diteruskan ke dalam layer - layer tersembunyi hingga menghasilkan nilai output di layer output. Fase kedua yaitu akan dihitung nilai error dari hasil output jaringan dengan output yang sebenarnya. Kemudian hasil tersebut dikembalikan ke layer tersembunyi hingga kembali ke layer input dimana saat proses pengembalian ini,

bobot nilai akan diubah untuk agar mendapatkan hasil output yang lebih akurat (Prasetyo, 2012).

Backpropagation Klasifikasi memiliki metode yang digunakan untuk melakukan pembelajaran terhadap kumpulan data dan kemudian memetakan masing-masing data yang terpilih ke dalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya. Tujuan dari klasifikasi yaitu memperkirakan kelas yang dimiliki dari suatu objek dimana objek tersebut belum diketahui labelnya. Proses klasifikasi ini melakukan proses pencarian model atau fungsi yang dapat menjelaskan atau membedakan kelas dari data tertentu (Han, et al., 2012). Untuk mengukur kualitas dari klasifikasi dapat digunakan perhitungan berikut ini:

2.2.3. Temporal Data Mining

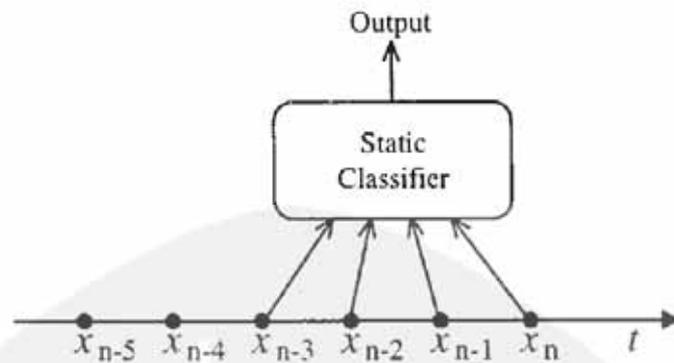
Temporal Data mining adalah sebuah teknik penambangan data yang dilakukan pada database temporal. Database temporal merupakan database yang memiliki data bersifat temporal yang artinya memiliki faktor waktu. Sebuah data pada temporal database bersifat *time series* yang artinya nilai pada data tersebut tersusun berdasarkan waktu tertentu. Dengan kata lain, *temporal data mining* berarti melakukan analisis penambangan data yang mempertimbangkan pengaruh waktu (Mitsa, 2010). Pada penambangan data ini, unsur kompleksitas akan muncul dikarenakan penggalian data dan pemrosesan data dilakukan berdasarkan runtun waktu tertentu. Dalam penerapannya, penambangan data temporal ini dapat digunakan untuk melakukan klasifikasi serta prediksi suatu data yang akan datang. Klasifikasi pada

temporal data mining dilakukan berdasarkan analisis trend dan pola tertentu berdasarkan waktu (Shahnawaz, et al., 2011). Klasifikasi dan prediksi dalam penambangan data temporal ini berdasarkan pergerakan nilai data pada suatu interval waktu yang diketahui bahwa pola-pola data yang muncul pada masa lalu dapat terulang.

Perbedaan utama antara klasifikasi statis yang biasanya dilakukan dengan klasifikasi pada data temporal sangat jelas terlihat pada dimensi waktu. Pada klasifikasi statis, tidak ada hubungan antara satu data dengan data yang lain, sehingga data dapat diproses secara satu per satu atau individual. Berbeda dengan klasifikasi yang dilakukan pada data temporal, proses dilakukan pada sebuah set data yang berurutan dari kumpulan data. Dalam proses klasifikasi temporal data ini, terdapat 2 jenis pendekatan untuk melakukan pemrosesan data pada temporal data, yaitu *External Temporal Processing* dan *Internal Temporal Processing* (Haselstainer & Pfurtscheller, 2000).

a. *External Temporal Processing*

External Temporal Processing memiliki arti bahwa pengolahan data temporal dimana dimensi waktu dari kumpulan data tersebut ditangani di luar proses klasifikasi. Gambar 1 di bawah ini dapat menunjukkan proses dari *External Temporal Processing*.



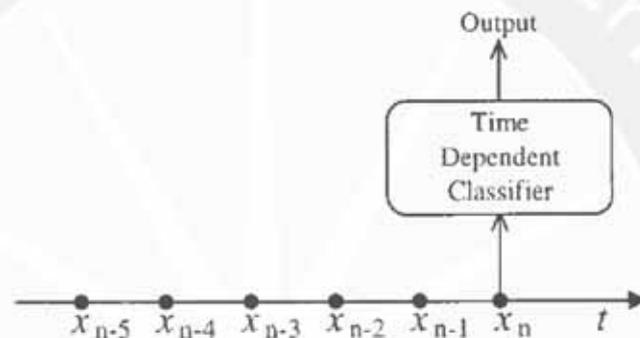
Gambar 2.2. External Temporal Processing (Haselstainer & Pfurtscheller, 2000).

Dari gambar 1 di atas, dapat dilihat bahwa proses ini membentuk sebuah pola yang digunakan sebagai inputan untuk klasifikasi menggunakan sebuah subset yang terdiri dari data-data individual. Singkatnya, kumpulan data individual akan dibentuk ulang menjadi sebuah pola yang lebih besar. Dari pola yang terbentuk tersebut akan merepresentasikan bentuk baru dari kumpulan data temporal yang ada sebelumnya. Pola baru dari kumpulan data individual tersebut dapat dibentuk dengan menggunakan perhitungan statistika.

Keuntungan utama dari pendekatan ini yaitu proses klasifikasi dapat dilakukan dengan menggunakan model klasifikasi yang sederhana yaitu klasifikasi untuk pola yang statis. Proses yang kompleks akan terdapat pada proses pembentukan pola baru dari kumpulan data individual.

b. Internal Temporal Processing

Pada pendekatan ini, dimensi waktu akan dilakukan bersamaan dengan proses klasifikasi. Sehingga satu buah data individual dari kumpulan data temporal juga akan menjadi satu inputan untuk diklasifikasikan. Gambar 2 di bawah ini dapat menunjukkan proses dari *Internal Temporal Processing*.



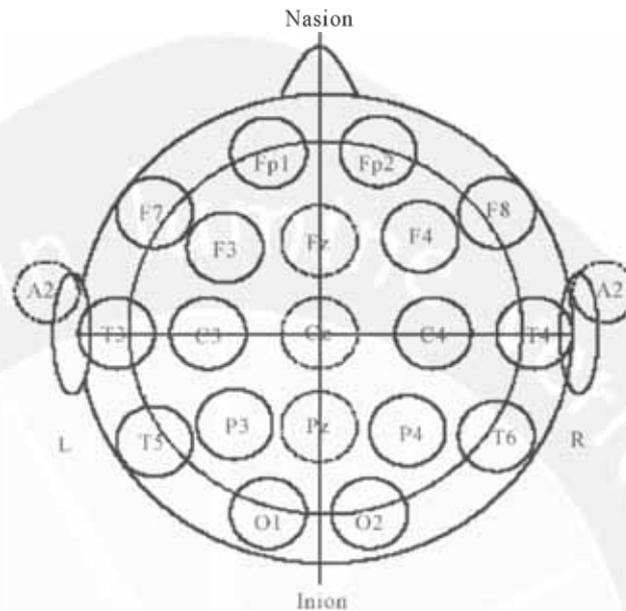
Gambar 2.3. Internal Temporal Processing (Haselstainer & Pfurtscheller, 2000).

Pada suatu waktu tertentu hanya akan ada satu data dari kumpulan data temporal yang digunakan sebagai satu inputan dalam proses klasifikasi. Di dalam proses klasifikasi juga akan dilakukan proses pengolahan data dengan metode-metode tertentu untuk data yang akan diproses.

2.2.4. Electroencephalography (EEG)

Electroencephalography atau EEG merupakan teknik untuk merekam aktifitas listrik yang ada di otak dimana hasil perekaman EEG merupakan sinyal analog (Andrzejak et al., 2001). Teknik penempatan electrode pada saat perekaman EEG yang biasa

digunakan adalah system international 10-20. Berikut ini letak elktrode dengan system internasional 10-20.



Gambar 2.4. Penempatan Elektrode dengan Sistem Internasional 10-20. (Song, 2011)

Penempatan elektroda pada gambar 2.4 diatas adalah penempatan elektroda dipermukaan kepala. Hasil dari perekaman aktifitas listrik dari penempatan electrode ini akan menggambarkan kondisi gelombang yang ada pada otak.

Terdapat 4 gelombang yang akan terekam dalam EEG. Yang pertama adalah gelombang alfa, beta, theta, dan delta serta gamma(Nandish et al., 2012). Gelombang alfa adalah gelombang dengan frekuensi 8-13 Hz yang akan muncul pada saat seseorang sedang mengalami kondisi rileks atau mata tertutup. Gelombang alfa dapat digunakan untuk melihat kenormalan suatu fungsi atau kerja otak. Gelombang beta adalah gelombang yang memiliki frekuensi 13 Hz

hingga 30 Hz yang timbul ketika orang sedang berpikir atau melakukan aktifitas. Gelombang theta adalah gelombang dengan frekuensi 4 sampai 8 Hz. Gelombang theta muncul pada saat seseorang tidur ringan atau dalam keadaan senang. Sedangkan gelombang delta adalah gelombang yang berfrekuensi rendah yaitu sekitar 0 sampai 4 Hz yang biasanya muncul pada saat seseorang tidur nyenyak. Gelombang ini dapat digunakan untuk mengidentifikasi kondisi *pathologic* seseorang seperti terjadinya luka atau infeksi, serta penyakit seperti kanker, tumor, dan epilepsi. Terakhir adalah gelombang gamma. Gelombang ini berfrekuensi lebih dari 30 Hz yang berkaitan dengan aktifitas otak untuk mengintegrasikan bermacam rangsangan.

Data hasil perekaman EEG akan diolah menggunakan komputer untuk mengetahui hasil atau kondisi seseorang. Hasil mentah dari EEG adalah sinyal analog, sedangkan pada komputer hanya dapat membaca data digital. Maka dari itu dibutuhkan perangkat *Analog to Digital Converter* (ADC) yang dapat mengubah sinyal analog kontinyu dari EEG menjadi sinyal digital yang dapat dibaca oleh komputer.

2.2.5. Principle Component Analysis dan Independent Component Analysis

Metode statistic yang sering digunakan untuk pengolahan sinyal adalah *Principal Component Analysis* (PCA) dan *Independent Component Analysis* (ICA). PCA sebuah metode yang dapat digunakan untuk mengekstraksi informasi atau parameter dari kumpulan data yang besar seperti data hasil perekaman sinyal

EEG. Parameter yang dapat digunakan pada PCA untuk mengekstraksi data EEG adalah rata-rata, varian, dan standar deviasi. *Independent Component Analysis* (ICA) merupakan metode yang digunakan untuk mencari informasi yang tersembunyi dari kumpulan data dimana informasi yang ditelusuri adalah faktor yang independen secara statistik dan non gaussian. Metode ini banyak digunakan untuk pemrosesan data sinyal seperti data EEG. Parameter yang sering digunakan untuk mengekstrak data sinyal EEG pada metode ini adalah *skewness* dan *kurtosis* (Cohen, 2014). Kedua parameter ini masuk dalam golongan *higher order statistic*. Dalam hal pengolahan sinyal, parameter *higher order statistic* lebih merepresentasikan data karena informasi yang penting lebih terdapat pada parameter ini. Kedua metode ini digunakan untuk menggambarkan dan merepresentasikan dua kondisi data apabila data sinyal terdistribusi Gaussian dan apabila data sinyal terdistribusi non-Gaussian. Berikut ini adalah parameter statistik yang digunakan untuk analisa sinyal EEG :

a. Rata-rata (*Mean*)

Mean merupakan nilai dari ukuran tendensi sentral yang didefinisikan sebagai nilai rata-rata dari kumpulan data. Nilai *mean* didapatkan dengan menghitung jumlah semua nilai dalam kumpulan data kemudian membagi penjumlahan data tersebut dengan banyaknya jumlah kumpulan data tersebut.

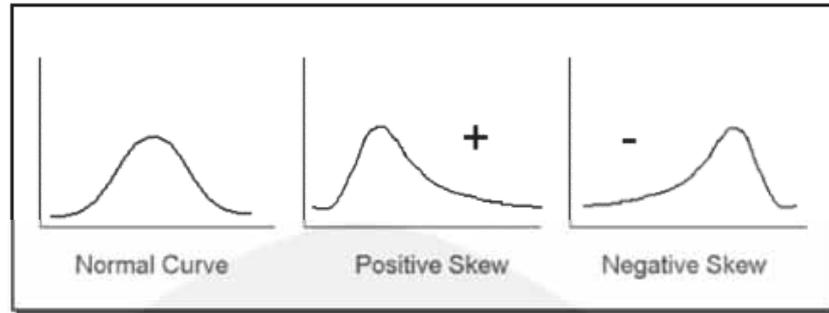
b. Standar Deviasi (*Standard Deviation*)

Standar Deviasi merupakan ukuran sebaran (dispersi) suatu distribusi. Jika suatu distribusi

rata maka nilai standar deviasinya kecil sedangkan jika suatu distribusi berfluktuasi maka nilai standar deviasinya besar. Standard deviasi adalah sebuah nilai yang menunjukkan seberapa banyak variasi yang terdapat pada mean atau nilai yang diharapkan. Standard deviasi dari sebuah variabel acak, populasi statistik, kumpulan data atau distribusi probabilitas adalah akar kuadrat dari varian. Nilai standard deviasi yang rendah menunjukkan nilai dari data-data yang ada itu dekat dengan nilai *mean* atau nilai yang diharapkan. Nilai standard deviasi yang tinggi menunjukkan bahwa nilai dari data-data itu jauh dari nilai mean atau nilai yang diharapkan. Standard deviasi secara umum dapat digunakan untuk mengetahui seberapa dekat semua variasi data dengan nilai *mean* dari data-data tersebut. Standard deviasi penting untuk menunjukkan seberapa besar perbedaan dari sekumpulan data yang ada dengan nilai *mean* atau nilai yang diharapkan.

c. Skewness

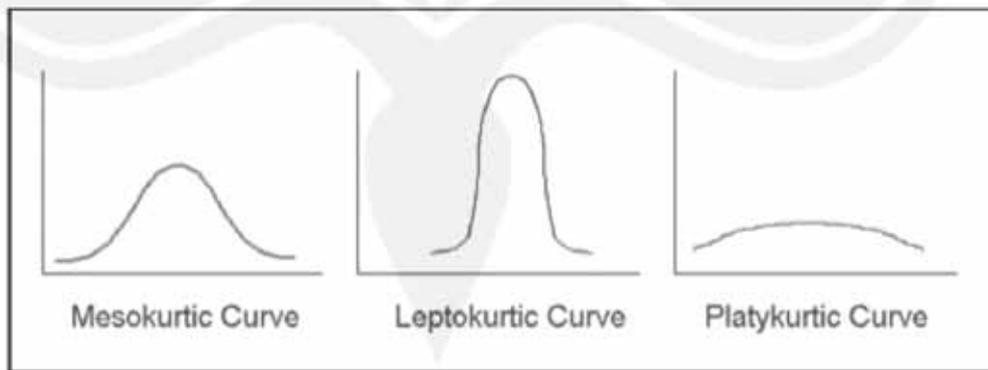
Skewness merupakan parameter ukuran yang mendefinisikan penyimpangan atau kemiringan dari kesimetrisan distribusi normal atau Gaussian pada sekumpulan data. Nilai *skewness* akan bernilai nol jika distribusi data simetris dengan *baseline* dan sebaliknya, nilai *skewness* bernilai tidak nol jika tidak simetris terhadap *baseline*. Nilai dari parameter ini akan bernilai positif jika ekor lebih condong ke kanan dan akan bernilai negatif jika ekor lebih condong ke kiri.



Gambar 2.5. Parameter Skewness

d. Kurtosis

Kurtosis adalah parameter ukuran yang mendefinisikan kepuncakan atau kedataran suatu distribusi pada kumpulan data. Terdapat 3 jenis distribusi yang ada pada parameter ini. Yang pertama yaitu *leptokurtic* yang akan terjadi jika nilai kurtosis positif atau lebih besar dari distribusi normal atau secara singkat apabila nilai kurtosis lebih berpuncak dibanding distribusi normal. Yang kedua adalah distribusi *mesokurtic* apa bila nilai *kurtosis* adalah nol dan jika nilai *kurtosis* bernilai negatif maka disebut sebagai distribusi *platykurtic*.



Gambar 2.6. Parameter Kurtosis

2.2.6. Waikato Environment for Knowledge Analysis (WEKA)

WEKA merupakan perangkat lunak yang menyediakan layanan untuk melakukan pengolahan data dalam *data mining*. Perangkat lunak ini berbasis *open source* dan dibuat menggunakan Java. WEKA dibuat dan dikembangkan oleh Universitas Waikato di Selandia Baru. WEKA merupakan perangkat lunak gratis yang tersedia dibawah *General Public License*. Perangkat ini memiliki fasilitas untuk melakukan *preprocessing data*, *classification*, *regression*, *clustering*, *association rules*, dan *visualiszation*.

Di dalam WEKA terdapat *library* kelas Java yang lengkap untuk melakukan implementasi metode mesin pembelajaran dan algoritma *data mining*. Kemudahan dari aplikasi WEKA ini adalah *library* kelas yang ada dapat digunakan secara bebas pada aplikasi lain. Bahkan, kelas yang ada pada WEKA dapat dijalankan pada aplikasi berbasis web. Hal ini memungkinkan pengguna untuk menerapkan berbagai teknik dan metode *data mining* pada weka untuk data yang ingin mereka olah sendiri tanpa memperhatikan *platform* komputer yang digunakan. (Witten, Frank & Hall, 2011)



Gambar 2.7. Tampilan Awal WEKA

WEKA mulai dikembangkan sejak tahun 1994 dan telah menjadi *software data mining open source* yang paling populer. WEKA mempunyai kelebihan seperti mempunyai banyak algoritma *data mining* dan *machine learning*, kemudahan dalam penggunaannya, selalu *up-to-date* dengan algoritma-algoritma yang baru. *Software* WEKA tidak hanya digunakan untuk akademik saja namun cukup banyak dipakai oleh perusahaan untuk meramalkan bisnis dari suatu perusahaan. Ian H. Witten merupakan latar belakang dibalik kesuksesan WEKA. Beliau merupakan profesor di *Universitas of Waikato, New Zealand*, yang menekuni *Digital Library, Text Mining, Machine Learning* dan *Information Retrieval*. Pada Weka ada beberapa metode pemilihan *variable* dari suatu *dataset*, diantaranya *BestFirst, ExhaustiveSearch, FCBFSearch, GeneticSearch, GreedyStepwise, RaceSearch, RandomSearch, Rankerdan, RankerSearch*. Metode atau Teknik yang digunakan Weka adalah *Predictive* dan *Descriptive* karena Weka mendukung teknik-teknik *data preprocessing, clustering, classification, regression, visualization*, dan *feature Reduction*. Semua teknik Weka adalah didasarkan pada asumsi bahwa data tersedia sebagai *flat file* tunggal atau hubungan, dimana setiap titik data digambarkan oleh sejumlah tetap atribut (biasanya, atribut numeric atau nominal, tetapi beberapa jenis atribut lain juga didukung). WEKA memiliki empat jenis *test option* yang dapat digunakan untuk melakukan proses klasifikasi. Proses yang ditangani adalah proses pelatihan dan pengujian. Keempat jenis *test option* tersebut yaitu:

a. *Use training set*

Klasifikasi ini akan menggunakan satu data untuk melakukan pelatihan. Kemudian seluruh data yang dilatih sebelumnya juga digunakan untuk proses pengujian.

b. *Supplied test set*

Klasifikasi ini dilakukan evaluasi dengan cara memprediksi seberapa baik satu dataset yang diambil dari sebuah data tertentu yang memang sudah disediakan untuk pengujian. Proses pelatihan akan dilakukan terlebih dahulu menggunakan data latih kemudian proses pengujian akan dilakukan dengan data uji yang berbeda dengan data yang dilatih pada klasifikasi.

c. *Cross-validation*

Klasifikasi ini dilakukan evaluasi dengan *cross-validation* dan menggunakan jumlah *fold* yang tertentu yang dapat diinputkan manual. Pada *cross-validation* akan ada pilihan berapa *fold* yang akan digunakan. Nilai *fold* awal yang diberikan adalah 10. Proses pengujian akan dilakukan sebanyak nilai *fold* yang diberikan serta akan dibentuk subset sebanyak nilai *fold*. Kemudian proses pengujian akan dilakukan menggunakan sebuah subset yang terbentuk dan sisanya akan digunakan untuk proses pelatihannya.

d. *Percentage split*

Klasifikasi ini dilakukan evaluasi dengan melakukan pembagian data antara data uji dan data latih pada satu dataset dengan menggunakan prosentase. Prosentase yang diinputkan akan

digunakan untuk proses pelatihan dan sisanya akan digunakan untuk proses pengujian. Proses ini biasanya dilakukan dengan perbandingan 2/3 data untuk pelatihan dan 1/3 data digunakan untuk proses pengujian.

Klasifikasi data pada Weka dapat menggunakan ANN dengan memanfaatkan algoritma Backpropagation. Pada Weka, ANN diberi nama *Multilayer Preception*. Jaringan yang ada pada Weka dapat dibuat dengan cara memberikan nilai-nilai pada parameter yang dibutuhkan oleh jaringan. Jaringan tersebut juga dapat pantau dan dimodifikasi selama proses pelatihan. Berikut ini parameter-parameter yang dapat digunakan untuk membangun sebuah jaringan :

a. *momentum*

Parameter ini digunakan untuk menyesuaikan bobot yang dimanfaatkan untuk menemukan nilai global eror yang minimal selama pelatihan. Parameter ini juga dapat mempercepat proses pelatihan. Nilai momentum dapat diberikan dari 0 hingga 1.

b. *hiddenLayers*

Parameter ini digunakan untuk memberi jumlah layer tersembunyi dari jaringan. Parameter ini mendukung pembuatan multilayer dengan menggunakan tanda koma untuk memisahkan jumlah node yang dibuat di tiap layer.

c. *trainingTime*

Parameter ini digunakan untuk memberi nilai epoch.

d. *learningRate*

Parameter ini digunakan untuk memberi nilai pada laju belajar yang ada pada jaringan.