

BAB 2 LANDASAN KEPUSTAKAAN

Pada bab ini akan dijelaskan mengenai kajian pustaka dan dasar teori yang digunakan dalam mendukung penelitian ini. Kajian pustaka berisi mengenai penelitian yang sudah dilakukan sebelumnya. Dasar teori berisi tentang teori yang digunakan sebagai bahan penelitian ini. Teori dasar yang digunakan sebagai bahan penelitian ini yaitu terdiri dari teori mengenai penyakit skizofrenia, kecerdasan buatan, *data mining*, dan algoritma *Support Vector Machine*.

2.1 Kajian Pustaka

Beberapa penelitian dengan menggunakan metode *Support Vector Machine* telah banyak dilakukan sebelumnya. Pada bagian kajian pustaka akan dijelaskan mengenai perbandingan dan hasil analisis dari setiap penelitian yang telah dilakukan sebelumnya mengenai penyakit Skizofrenia dan metode *Support Vector Machine*. Penelitian yang pertama adalah penelitian yang berjudul “Implementasi *Certainty Factor* Dalam Sistem Pakar Untuk Melakukan Diagnosa Penyakit Gangguan Jiwa”. Penelitian ini membahas mengenai sistem pakar untuk mendiagnosa penyakit gangguan jiwa berdasarkan gejala-gejala yang dialami pasien. Metode yang dipakai dalam penelitian ini adalah dengan menggunakan metode *Certainty Factor*. Hasil akurasi yang didapatkan dengan menggunakan metode ini adalah sebesar 93%. Data yang digunakan masih terbilang sedikit karena hanya menggunakan 15 data uji (Taufiq & Natarsyah, 2016).

Penelitian kedua merupakan penelitian mengenai diagnosa penyakit Skizofrenia yang berjudul “*Automatic Schizophrenia Discrimination on fNIRS by Using PCA and SVM*”. Penelitian ini menggunakan teknik *functional Near-Infrared spectroscopy* (fNIRS), teknik ini biasanya digunakan dalam dunia medis dalam mendiagnosa suatu penyakit. Sinyal dari fNIRS akan diekstrak menggunakan metode *Principal Component Analysis* (PCA). Hasil dari ekstrak akan digunakan untuk proses *training* dalam klasifikasi menggunakan metode *Support Vector Machine* (SVM). *Dataset* yang digunakan sebanyak 52 data, dan *kernel* yang digunakan sebanyak 3 *kernel*. Hasil akurasi yang didapatkan dari *kernel* linier adalah sebesar 90%, pada *kernel polynomial* didapatkan akurasi 93,33%, sedangkan pada *kernel* RBF didapatkan hasil akurasi sebesar 73,33% (Song, et al., 2016).

Penelitian ketiga adalah penelitian yang berjudul “Penerapan Metode *Support Vector Machine* Pada Diagnosa Hepatitis”. Pada penelitian ini peneliti ingin membangun suatu sistem yang dapat mendiagnosa penyakit hepatitis berdasarkan data dari hasil tes fungsi hati. Metode yang digunakan dalam permasalahan pada penelitian ini adalah metode SVM. Hasil akurasi maksimal yang didapatkan pada penelitian ini cukup tinggi yaitu sebesar 83% dengan menggunakan *kernel* linier, sedangkan menggunakan *kernel* RBF didapatkan hasil akurasi maksimal sebesar 96%. Data yang digunakan sebanyak 579 data yang dibagi menjadi 200 data latih yang dibagi menjadi 100 data positif dan 100 data negatif, dan 379 sebagai data uji yang diambil secara *random*. Metode dalam

melakukan *training* data pada penelitian ini adalah dengan menggunakan metode *Quadratic Programming* (Munawarah, et al., 2016).

Penelitian keempat adalah penelitian berjudul “Klasifikasi Kondisi Detak Jantung Berdasarkan Hasil Rekam Elektrokardiografi (EKG) Menggunakan Algoritma *Support Vector Machine* (SVM)”. Penelitian ini membahas mengenai klasifikasi detak jantung menggunakan metode SVM. Dalam pengklasifikasian detak jantung pada penelitian ini dibagi dalam 4 kelas. Dalam penelitian ini *kernel* yang digunakan adalah *kernel polynomial*. Hasil akurasi tertinggi yang didapatkan adalah sebesar 87,5%, sedangkan rata-rata akurasi tertinggi adalah 81,3%. Data yang digunakan sebanyak 140 data yang digunakan sebagai data latih dan data uji (Uswatun, 2016).

Pada referensi kelima berjudul “Penerapan Metode Klasifikasi *Support Vector Machine* Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang”. Penelitian ini membahas mengenai penentuan akreditasi sekolah dasar menggunakan metode klasifikasi *Support Vector Machine*. *Kernel* yang digunakan pada penelitian ini adalah *kernel Polynomial* dan RBF. Hasil akurasi yang dihasilkan dengan menggunakan data *training* adalah sebesar 100% untuk *kernel* RBF, dan 98,81% pada *kernel Polynomial*. Dalam pengujian menggunakan data *testing*, hasil akurasi yang didapatkan adalah sebesar 93,902% dengan menggunakan *kernel* RBF, dan 92,683% dengan *kernel polynomial*. Data yang digunakan dalam penelitian ini sebanyak 337 data *training* dan 82 data *testing* (Octaviani, et al., 2014).

Penelitian keenam yang berjudul “*Heart Disorder Detection Based on Computerized Iriodology using Support Vector Machine*”. Penelitian ini menjelaskan mengenai pembangunan sistem otomatis untuk mendeteksi gangguan jantung melalui iris mata. Dalam penelitian ini peneliti menggunakan metode *Principal Component Analysis* (PCA) dan *Support Vector Machine* (SVM). Metode PCA digunakan untuk mengekstrasi gambar iris mata yang nantinya hasil dari ekstrasi akan menjadi input dalam proses klasifikasi, sedangkan metode SVM digunakan untuk mengklasifikasikan data yang dibagi klasifikasikan kedalam 2 kelas. Data yang digunakan sebanyak 50 data latih dan 40 data uji. Hasil akurasi yang didapatkan pada penelitian ini sebesar 80 % (Permatasari, et al., 2016).

Terdapat beberapa perbedaan pada penelitian ini dengan penelitian yang telah dijabarkan sebelumnya. Pertama objek yang diteliti adalah penyakit kejiwaan Skizofrenia dengan menggunakan metode klasifikasi *Support Vector Machine* (SVM). Perbedaan kedua klasifikasi dibagi menjadi 5 kelas berdasarkan jenis-jenis penyakit skizofrenia yaitu *paranoid*, *hebefrenik*, katatonik, *undifferentiated*, simpleks. Lalu untuk menyelesaikan masalah *multiclass* pada penelitian ini, maka digunakan teknik *One Against All*. Perbedaan lainnya pada penelitian ini adalah akan menggunakan *K-fold Cross Validation* untuk melakukan proses pengujian.

Tabel 2.1 Kajian Pustaka

No	Judul	Objek (Input)	Metode	Hasil (Output)
1	Implementasi <i>Certainty Factor</i> Dalam Sistem Pakar Untuk Melakukan Diagnosa Penyakit Gangguan Jiwa	Objek: Penyakit Gangguan Jiwa	<i>Certainty Factor</i>	Memberikan hasil diagnosa dan kesesuaian hasil sistem dengan data pasien dari pakar. Hasil akurasi yang didapatkan adalah sebesar 93%
2	Penerapan Metode <i>Support Vector Machine</i> Pada Diagnosa Hepatitis	Objek: Penyakit Hepatitis	<i>Support Vector Machine</i> (SVM)	Hasil yang didapatkan adalah hasil diagnosa yang terbagi menjadi 2 kelas yaitu positif dan negatif. Hasil akurasi yang didapatkan adalah sebesar 83% dengan <i>kernel</i> linier dan 96% dengan <i>kernel</i> RBF
3	<i>Automatic Schizophrenia Discrimination on fNIRS by Using PCA and SVM</i>	Objek: Penyakit Skizofrenia	<i>Principal Component Analysis</i> (PCA) dan <i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini adalah hasil diagnosa penyakit Skizofrenia. Akurasi tertinggi yang didapatkan sebesar 93,33% pada <i>kernel polynomial</i>
4	Klasifikasi Kondisi Detak Jantung Berdasarkan Hasil Rekam Elektrokardiografi (EKG) Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	Objek: Kondisi Detak Jantung	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini adalah hasil diagnosa detak jantung yang terbagi menjadi 4 kelas. Akurasi tertinggi yang didapatkan sebesar 81,33% dengan menggunakan <i>kernel polynomial</i>
5	Penerapan Metode Klasifikasi <i>Support Vector Machine</i> Pada	Objek: Data Akreditasi Sekolah Dasar di Kabupaten	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini adalah status akreditasi sekolah dasar yang terbaru. Akurasi tertinggi yang

	Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang	Magelang		didapatkan adalah sebesar 93,902% pada <i>kernel</i> RBF
6	<i>Heart Disorder Detection Based on Computerized Iriodology using Support Vector Machine</i>	Objek: Penyakit Hati	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini adalah hasil diagnosa penyakit hati, dan akurasi yang didapatkan adalah sebesar 80%
7	Penerapan Algoritma <i>Support Vector Machine</i> (SVM) Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia	Objek: Penyakit Kejiwaan Skizofrenia	<i>Support Vector Machine</i> (SVM)	Hasil dari penelitian ini adalah hasil diagnosa jenis penyakit skizofrenia, akurasi tertinggi yang didapatkan adalah sebesar 59,09% dengan menggunakan <i>kernel polynomial</i>

Sumber: (Taufiq & Natarsyah, 2016), (Munawarah, et al., 2016), (Song , et al., 2016), (Uswatun, 2016), (Octaviani, et al., 2014), (Permatasari, et al., 2016)

2.2 Skizofrenia

Skizofrenia merupakan suatu gangguan terhadap fungsi otak yang merusak cara berpikir seseorang, sehingga tidak terjadi keselarasan antara pola berpikir dengan tingka laku orang tersebut (Hendarsyah, 2016). Salah satu akibat dari penyakit ini adalah penderita tidak dapat melakukan kegiatan sosial dengan baik dikarenakan adanya perbedaan tingkah laku pasien dengan tingkah laku manusia seperti biasa. Pada penderita Skizofrenia yang dalam keadaan kritis biasanya tidak dapat lagi untuk melakukan kegiatan secara mandiri, contohnya dalam hal menjaga kebersihan pasien memerlukan bantuan dari *volunteer*. Upaya dalam mengobati penyakit Skizofrenia agar dapat kembali berpikir dan berperilaku seperti manusia normal adalah dengan cara mengikuti rehabilitasi (Maryatun, 2015). Beberapa terapi dilakukan pada program rehabilitasi, contohnya seperti terapi gerak. Pada terapi ini pasien akan melakukan olahraga, dimana kegiatan olahraga ini akan dapat memperbaiki sel *neuron* yang rusak pada otak.

2.2.1 Gejala Skizofrenia

Gejala pada penyakit Skizofrenia dikategorikan berdasarkan dampak yang ditimbulkan. Berdasarkan buku pedoman diagnosis yang dipakai oleh para dokter spesialis kejiwaan yaitu PPDGJ jilid 3 (Maslim, 2001), dalam mendiagnosis

penyakit skizofrenia pasien dapat dikatakan mengalami penyakit skizofrenia jika mengalami minimal satu gejala yang sangat jelas seperti:

a. Isi Pikiran

1. *Thought Echo* (Sisip Pikir)

Gejala ini merupakan gejala dimana penderita akan mendengar isi pikirannya sendiri seperti ada yang membisikkan ke telinganya.

2. *Thought Insertion or Withdrawal* (Sedot Pikir)

Gejala ini juga merupakan gejala dimana penderita tidak dapat mengingat isi pikirannya, karena penderita akan merasa isi pikirannya telah dicuri oleh orang lain.

3. *Thought Broadcasting* (Siar Pikir)

Pada gejala ini penderita akan merasa isi pikirannya tersiar keluar, sehingga penderita merasa masyarakat disekitarnya tau isi pikirannya.

b. Delusi

Delusi adalah suatu gejala yang dialami oleh penderita dimana penderita akan mempercayai suatu kebenaran yang sebenarnya tidak mungkin terjadi (Wiramihardja, 2015). Salah satu contohnya seperti mempercayai bahwa manusia dapat terbang atau dapat berjalan diatas air. Seorang yang mengalami delusi akan dikuasai oleh kepercayaan yang dianggapnya benar. Mereka tidak akan merasa bahwa yang percayai benar ada salahnya, karena mereka akan selalu mencari pembenaran pada opini mereka. Menurut Wiramihardja (2015) delusi dibagi menjadi 4 tipe yaitu sebagai berikut:

1. Delusi penyiksaan

Delusi ini merupakan delusi dimana penderita meyakini bahwa ada seseorang yang melakukan kekerasan atau penyiksaan pada dirinya, contohnya penderita meyakini bahwa beberapa agen kepolisian akan menangkap dirinya karena suatu kejadian yang menimpanya.

2. Delusi grandiose

Delusi ini merupakan keyakinan pada penderita bahwa dirinya adalah seseorang yang memiliki bakat, pengetahuan, atau kekuatan yang diluar imajinasi, contohnya penderita ini meyakini bahwa dirinya adalah seorang utusan dari suatu kerajaan yang sudah tidak ada.

3. Delusi rujukan

Delusi ini merupakan salah satu keyakinan pada penderita bahwa setiap gerak-geriknya diawasi oleh suatu organisasi.

4. Delusi diawasi

Delusi ini merupakan delusi dimana penderita merasa bahwa dirinya telah dikendalikan oleh kekuatan yang diluar jangkauan, contohnya dia

merasa bahwa ada sosok makhluk luar angkasa yang mengendalikan seluruh anggota badannya.

c. Halusinasi

Halusinasi merupakan gejala dimana seseorang mengalami penglihatan terhadap suatu objek yang sebenarnya tidak ada (Wiramihardja, 2015). Halusinasi juga dapat dialami oleh orang yang tidak teridentifikasi mengalami penyakit Skizofrenia, tetapi perbedaannya adalah penderita Skizofrenia mengalami halusinasi setiap saat, bukan hanya sesaat karena faktor kelelahan atau faktor pengaruh minuman beralkohol seperti yang dialami orang pada umumnya. Terdapat beberapa jenis halusinasi yang biasanya dialami penderita Skizofrenia, contohnya seperti:

1. Halusinasi Pendengaran

Halusinasi ini terjadi ketika seorang merasakan suatu bunyi ataupun suara yang sebenarnya tidak ada.

2. Halusinasi Perabaan

Halusinasi ini terjadi ketika seorang merasakan ada yang menyentuh tubuhnya seperti binatang merayap ataupun serangga, yang sebenarnya tidak ada sama sekali binatang pada tubuhnya.

3. Halusinasi Somatis

Halusinasi ini terjadi ketika seorang merasakan ada sesuatu yang terjadi di dalam tubuhnya, contohnya seperti ada ulat yang sedang memakan salah satu organ tubuhnya.

d. Waham

1. Waham Pikiran dan Pembicaraan

Seorang penderita Skizofrenia biasanya akan mengalami disorganisasi terhadap cara berpikir dengan pembicaraan yang dilakukannya. Salah satu ciri disorganisasi yang dialami penderita Skizofrenia yaitu, penderita tidak dapat menjawab pertanyaan secara jelas, dan penderita Skizofrenia biasanya akan menjawab pertanyaan yang sebenarnya bukan topik dari pertanyaan tersebut (Wiramihardja, 2015).

2. Waham Perilaku atau Tingkah Laku Katatonik

Penderita Skizofrenia sendiri terkadang melakukan hal-hal yang tidak bisa diperkirakan. Salah satu contohnya, terkadang penderita Skizofrenia akan berteriak secara tiba-tiba. Perilaku yang dilakukan oleh penderita Skizofrenia bisa dikatakan bahwa semua yang dilakukan tidak wajar, tidak seperti yang dilakukan oleh manusia pada umumnya.

Adapun cara lain dalam mendiagnosis penyakit Skizofrenia yaitu, pasien dapat dikatakan mengalami penyakit Skizofrenia jika mengalami minimal dua gejala yang sangat jelas seperti:

a) *Affective Flattening*

Affective Flattening adalah gejala dimana penderita akan mengalami penurunan atau berkurangnya reaksi terhadap rasa emosional mereka. Pada umumnya gejala ini menyebabkan penderita akan memberikan ekspresi yang datar terhadap semua jenis kejadian baik itu kejadian yang menyedihkan maupun kejadian yang menyenangkan. Dalam beberapa kejadian, seorang penderita skizofrenia tidak dapat mengekspresikan rasa emosional dalam hatinya, walaupun dalam hatinya terjadi suatu reaksi emosional yang sangat tinggi.

b) *Alogia*

Alogia merupakan salah satu gejala dimana penderita skizofrenia mengalami penurunan terhadap cara berbicara. Pada gejala ini penderita akan menjawab pertanyaan dari siapapun secara singkat dan tidak jelas. Hal ini menyebabkan penderita akan mengurangi kegiatan sosial terutama melakukan percakapan kepada orang lain.

c) *Avolition*

Avolition adalah gejala pada penderita Skizofrenia yang menyebabkan terjadinya pengurangan terhadap motivasi melakukan kegiatan sehari-hari. Salah satu contohnya penderita gejala ini akan menghabiskan waktu hanya untuk berdiam diri dan merenung.

2.2.2 Jenis-jenis Skizofrenia

Penyakit Skizofrenia diklasifikasikan dalam 5 jenis yang dikategorikan berdasarkan gejala-gejala yang dialami oleh pasien. Berikut ini adalah jenis-jenis dari penyakit skizofrenia (Wiramihardja, 2015):

a. Tipe *Paranoid*

Seseorang penderita skizofrenia tipe *paranoid* memiliki beberapa tingkah laku yang berbahaya. Hal ini disebabkan penderita skizofrenia tipe *paranoid* mengalami halusinasi yang kebanyakan bertema penyiksaan. Para penderita ini juga biasanya lebih pandai dalam mengutarakan pendapat mengenai suatu komplotan yang akan melawan dirinya. Mereka akan melawan semua argumen yang berbeda dengan delusi yang ada dalam pikirannya, sehingga mereka dapat melakukan tindakan yang arogan dan berbahaya. Dalam beberapa kasus, penderita tipe ini dapat mengemukakan isi hatinya yang sakit diakibatkan penyiksaan yang telah dilakukan kepada dirinya.

b. Tipe *Hebrefrenik*

Skizofrenia tipe ini merupakan jenis skizofrenia yang biasanya menimpa orang pada usia muda dan juga paling pertama bila dibandingkan antara gejala-gejala skizofrenia lainnya. Tipe ini juga biasanya disebut tipe *hebrefrenik*. Beberapa perilaku yang merupakan ciri dari penderita skizofrenia tipe disorganisasi adalah sebagai berikut:

1. Penderita akan sering tersenyum yang tidak wajar dan bisa juga tertawa keras di keadaan yang sebenarnya tidak lucu sama sekali.
2. Delusi yang berlebihan mengenai banyak hal mulai dari hal yang berkaitan dengan seksual, penyiksaan, religius, dan fantasi-fantasi yang lain.
3. Penderita akan berbicara layaknya bayi, tertawa yang kekanak-kanakan, dan juga sering mengulang bunyi-bunyi tertentu.
4. Mereka tidak dapat melakukan kegiatan seperti mandi, makan, dan juga berpakaian secara mandiri.

Penderita sering kali mengutarakan ekspresi yang tidak sesuai dengan pembicaraan yang diutarakannya.

c. Tipe Katatonik

Para penderita Skizofrenia tipe katatonik biasanya tidak bisa lagi bersosialisasi dengan baik, hal ini dikarenakan terjadinya *withdrwal* (penarikan diri) terhadap lingkungan disekitarnya. Mereka cenderung menjadi orang yang sangat individual. Terkadang mereka hanya duduk diam dengan posisi yang sama dalam jangka waktu yang lama. Beberapa perilaku yang biasanya digunakan untuk mendiagnosis penderita skizofrenia tipe katatonik adalah sebagai berikut:

1. Penderita tidak bergerak dalam jangka waktu yang cukup panjang.
2. Penderita merasa bahagia secara tiba-tiba yang diekspresikan secara berlebihan.
3. Melakukan hal-hal yang aneh seperti sering bertepuk tangan atau bergemeringsing.
4. Sering kali mengucapkan kata-kata yang didengar dari orang lain dan diucapkan berulang-ulang.

d. Tipe *Undifferentiated*

Seseorang yang mengalami gejala skizofrenia tetapi tidak dapat dikategorikan ke tipe manapun akan termasuk pada jenis skizofrenia tipe *undifferentiated*. Mereka tidak dapat dikategorikan karena terjadinya perubahan perilaku yang begitu cepat. Beberapa gejala yang biasanya dialami oleh penderita ini adalah seperti:

1. Memiliki pandangan yang mudah berubah
2. Mengalami delusi
3. Terkadang mengalami ketakutan yang tidak wajar
4. Mengalami depresi
5. Emosi yang cepat berubah

e. Tipe Simpleks

Tipe Skizofrenia simpleks adalah salah satu jenis skizofrenia yang cukup sulit didiagnosis, karena dalam menentukan tipe ini bergantung pada proses

perkembangan yang sedang berjalan pada diri pasien (Maslim, 2001). Proses yang dilihat adalah perkembangan gejala negatif pada pasien, yang terjadi tanpa ada catatan sebelumnya mengenai gejala negatif yang dialaminya.

Perbedaan yang mencolok dari tipe skizofrenia simpleks dengan tipe skizofrenia lainnya adalah psikotik yang tidak jelas. Psikotik merupakan kemampuan untuk menilai kenyataan pada dirinya.

2.3 Kecerdasan Buatan

Kecerdasan buatan (*Artificial Intelligence*) merupakan suatu cabang ilmu pengetahuan yang menggunakan sistem yang dapat bekerja seperti otak manusia untuk memecahkan suatu masalah yang rumit (Nugraha & Winiarti, 2014). Pengertian lain dari kecerdasan buatan menurut Nasution (2012) adalah pengetahuan yang didapatkan dari suatu *entitas* buatan. Cara kerja sistem ini didapatkan dari cara pola berpikir manusia dan diimplementasikan ke bahasa mesin pada komputer.

Tujuan utama dari kecerdasan buatan adalah membuat suatu sistem yang dapat membantu dan mempermudah pekerjaan manusia dalam kehidupan sehari-hari. Permasalahan yang dapat diselesaikan oleh mesin kecerdasan buatan adalah seperti diagnosa penyakit, memprediksi cuaca, hingga membantu dalam kegiatan industri. Kekurangan yang dimiliki manusia dalam bekerja seperti kurang teliti maupun keterbatasan fisik bisa dibantu dengan menggunakan teknologi dari kecerdasan buatan. Beberapa ruang lingkup pada kecerdasan buatan yaitu, sistem pakar, *computer vision*, robotika, dan yang lainnya.

2.4 Data Mining

Data Mining merupakan teknik atau cara dalam menemukan suatu informasi tersembunyi dari suatu *dataset* yang banyak (Almasoud, et al., 2015). Informasi dari suatu *dataset* didapatkan dengan cara mempelajari pola-pola, hubungan, maupun aturan yang terdapat pada *dataset*. Definisi lain mengenai *data mining* adalah suatu analisis pada *database* untuk menemukan pola dan hubungan pada sekelompok data (PETRE & Ruxandra, 2012). *Data mining* adalah salah satu cabang keilmuan pada komputer yang saat ini sedang berkembang. Dalam *data mining* terdapat beberapa cabang ilmu yang diadopsi dalam penerapannya, contohnya statistik, kecerdasan buatan, dan pemodelan matematika. Di beberapa jurnal maupun karya ilmiah *data mining* sering dikenal dengan nama lain yaitu *Knowledge Discovery in Database* (KDD). Kedua istilah ini memiliki definisi yang sama yaitu menganalisis dan mengolah *database* untuk menghasilkan suatu data maupun informasi yang berguna.

2.4.1 Proses Data mining

Menurut Silwattananusarn & Tuamsuk (2012) dijelaskan terdapat beberapa tahapan dalam melakukan proses *data mining*, berikut ini merupakan tahap-tahap dalam melakukan *data mining*:

1. Seleksi

Seleksi merupakan tahapan dimana data akan di seleksi berdasarkan relevansi terhadap permasalahan untuk dianalisis dan diproses selanjutnya

2. *Preprocessing*

Tahapan ini merupakan tahap dimana data yang mengalami inkonsistensi akan dihapus dan juga menghilangkan *noise* pada data.

3. Transformasi

Tahap transformasi berfungsi sebagai tahap dimana data yang sudah di *preprocessing* akan di rubah formatnya sesuai teknik *data mining* yang digunakan

4. *Data Mining*

Terdapat beberapa algoritma yang dapat dipakai dalam melakukan *data mining*. Algoritma ini dibutuhkan untuk membentuk pola-pola pada data.

5. Evaluasi / Interpretasi

Pola-pola yang telah terbentuk akan dievaluasi, jika pola sudah memberikan hasil yang sesuai dengan hipotesa maka pola itu akan diinterpretasikan ke dalam bahasa yang dapat dimengerti manusia.

2.4.2 Klasifikasi

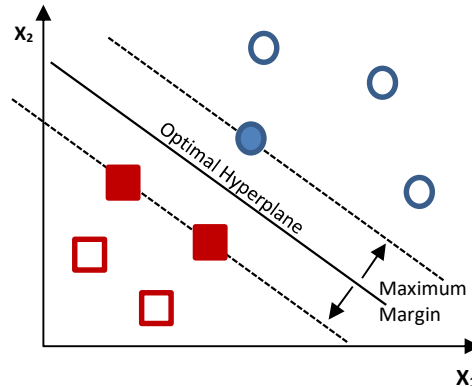
Definisi dari klasifikasi adalah suatu teknik pembelajaran dengan melakukan pelatihan pada suatu data untuk dikelompokkan berdasarkan vektor (fitur) ke dalam suatu kelas yang telah tersedia (Prasetyo, 2014). Klasifikasi juga dapat difinisikan suatu model pembelajaran dengan cara mencari pola suatu data untuk dimasukan ke beberapa kelas yang telah terdefiniskan. Dalam metode prediktif pada *data mining*, klasifikasi merupakan salah satu metode yang paling sering digunakan dalam menyelesaikan suatu masalah. Permasalahan yang biasanya menggunakan metode klasifikasi adalah seperti diagnosa penyakit, penentuan nilai, dan lainnya. Terdapat beberapa jenis algoritma yang dapat digunakan dalam teknik klasifikasi, contohnya adalah *K-Nearest Neighbor*, *Support Vector Machine*, *Artificial Neural Network*, *Decision Tree*, dan metode modifikasi lainnya.

2.5 *Support Vector Machine (SVM)*

Salah satu metode klasifikasi yang sedang banyak dikembangkan adalah algoritma *Support Vector Machine (SVM)*. Algoritma ini hanya menggunakan beberapa data latih untuk dijadikan sebagai proses klasifikasi. Hal ini yang menjadikan suatu kelebihan pada metode SVM dibandingkan metode klasifikasi lainnya yang menggunakan semua data latih untuk diproses (Prasetyo, 2014).

Konsep dasar dari algoritma SVM adalah dengan mencari *hyperplane* yang memiliki nilai terbaik untuk digunakan sebagai batas pemisah data. Cara untuk

mencari *hyperplane* terbaik adalah dengan menghitung atau mengukur *margin hyperplane*. Dalam algoritma SVM, *margin* merupakan jarak antara *hyperplane* ke data terdekat dari setiap kelas yang ada. Data terdekat yang menjadi nilai *margin hyperplane* disebut *Support Vector*.



Gambar 2.1 Garis *Hyperplane*

Sumber: (Prasetyo, 2014)

Metode SVM merupakan metode klasifikasi yang menggunakan garis *hyperplane* untuk mencari jarak diantara data pada kelas yang berbeda. Pada Gambar 2.1 dijelaskan bahwa garis *hyperplane* yang optimal akan memberikan jarak yang terbaik dalam memisahkan dua kelas data. Data pada suatu dataset diberikan variabel x_i , sedangkan untuk kelas pada dataset diberikan variabel y_i . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh *hyperplane* bernilai 1, sedangkan kelas lainnya bernilai -1. Maka persamaan yang didapatkan seperti Persamaan 2.1 dan Persamaan 2.2.

$$(w \cdot x_i + b) \geq 1 \text{ untuk } y_i = 1 \quad (2.1)$$

$$(w \cdot x_i + b) \leq -1 \text{ untuk } y_i = -1 \quad (2.2)$$

Keterangan:

x_i = Data ke- i

w = Nilai bobot *support vector*

b = Nilai bias

y_i = Kelas data ke- i

Bobot vektor (w) adalah garis vektor yang tegak lurus antara titik pusat koordinat dengan garis *hyperplane*. Bias (b) merupakan kordinat garis relative terhadap titik kordinat. Persamaan 2.3 merupakan persamaan untuk menghitung nilai b , sedangkan Persamaan 2.4 merupakan persamaan untuk mencari nilai w .

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (2.3)$$

$$w = \sum_{i=1}^n \alpha_i x_i y_i \quad (2.4)$$

Keterangan:

b = nilai bias
 $w \cdot x^+$ = nilai bobot untuk kelas data positif
 $w \cdot x^-$ = nilai bobot untuk kelas data negatif
 w = bobot vektor
 α_i = nilai *alpha* ke- i
 y_i = kelas data ke- i
 $K(x_i, x)$ = nilai *kernel* ke- i

Variabel α_i merupakan bobot tiap data yang dimana jika bernilai positif maka akan disebut *support vector*. Dalam proses pengujian untuk mendapatkan hasil klasifikasi pada setiap data uji, Persamaan 2.5 dan Persamaan 2.6 dapat digunakan.

$$f(x) = (w \cdot x + b) \text{ atau } f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (2.5)$$

Keterangan:

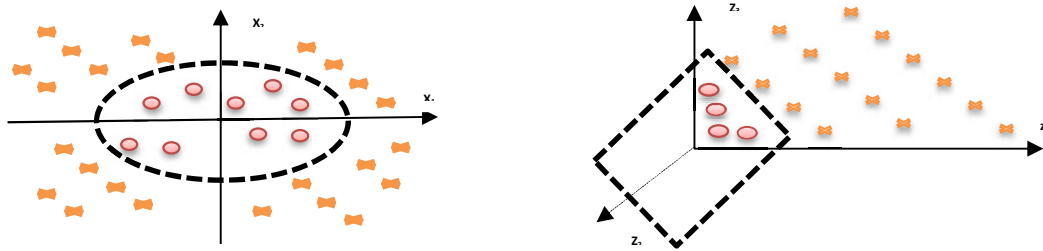
α_i = nilai alfa ke- i
 y_i = kelas data ke- i
 m = jumlah data
 $K(x_i, x)$ = *kernel*
 b = nilai bias

$$\text{Fungsi Klasifikasi} = \text{sign}(f(x)) \quad (2.6)$$

Jika hasil fungsi klasifikasi bernilai positif maka nilai dari $\text{sign}(f(x)) = 1$, dan jika hasil fungsi klasifikasi bernilai negatif maka nilai dari $\text{sign}(f(x)) = -1$. Jadi jika nilai hasil fungsi bernilai 1 maka data uji tersebut merupakan data kelas 1, dan sebaliknya jika bernilai -1 maka data uji tersebut merupakan data kelas 2.

2.5.1 Kernel Trick

Pada dasarnya metode SVM merupakan metode yang menggunakan *hyperplane* linier yang digunakan untuk memisahkan data secara linier. Dalam mengatasi permasalahan dimana data berupa data nonlinier, maka dapat digunakan teknik *kernel trick*. *Kernel* merupakan suatu fungsi yang digunakan untuk memetakan data dari dimensi awal ke dimensi yang lebih tinggi (Prasetyo, 2014). Pemetaan ini berfungsi untuk membuat suatu ruang vektor baru dimana data dapat diselesaikan secara linier.



Gambar 2.2 SVM Linier dan SVM NonLinier

Sumber : (Prasetyo, 2014)

Terdapat beberapa fungsi *kernel* yang dapat dipakai dalam memecahkan masalah SVM nonlinier:

1. Linier, ditunjukkan pada Persamaan 2.7

$$K(x, y) = x \cdot y \quad (2.7)$$

2. *Polynomial*, ditunjukkan pada Persamaan 2.8

$$K(x, y) = (x \cdot y)^2 \quad (2.8)$$

3. *Gaussian* RBF, ditunjukkan pada Persamaan 2.9

$$K(x, y) = \exp \frac{-||x-y||^2}{2\sigma^2} \quad (2.9)$$

Keterangan:

$K(x, y)$ = *kernel*

x = data

y = kelas data

σ = konstanta sigma

2.5.2 Sequential Training SVM

Sequential Training SVM adalah satu dari beberapa macam algoritma yang digunakan dalam melakukan proses *training* pada metode SVM. Selain dari *Sequential Training SVM* terdapat algoritma lainnya yaitu seperti *Sequential Minimal Optimazation* (SMO) dan *Quadratic Programming*. Kelebihan yang dimiliki dari algoritma *Sequential Training* adalah jika dibandingkan dengan algoritma laiinya yang cukup kompleks dan membutuhkan waktu yang lama, algoritma ini terbilang sederhana dalam proses pengerjaanya, dan waktu yang dibutuhkan dalam pengerjaanya lebih cepat. Persamaan 2.10 sampai 2.13 adalah algoritma dari *Sequential Training SVM*.

1. Inisialisasi nilai α_i dan konstanta $C, \lambda, \gamma, \varepsilon$.

Keterangan:

α_i = nilai untuk mendapatkan nilai *support vector*

C = nilai untuk membatasi nilai alfa

λ = nilai jarak *margin*

γ = nilai *learning rate*

ε = nilai *epsilon*

2. Perhitungan *matriks hessian* dengan menggunakan Persamaan 2.10.

$$D_{ij} = y_i y_j (K(x_i x_j) + \lambda^2) \quad \text{untuk } i, j = 1, \dots, n \quad (2.10)$$

Keterangan:

x_i = data ke-i

x_j = data ke-j

y_i = kelas data ke-i

y_j = kelas data ke-j

n = jumlah data

$K(x_i, x_j)$ = fungsi *kernel* yang digunakan

3. Tahap 3 dilakukan secara berulang sampai didapatkan keadaan iterasi maksimum (*itermax*) tercapai atau $\max(|\delta\alpha_i|) < \varepsilon$ (*epsilon*).

$$E_i = \sum_{j=1}^1 \alpha_i D_{ij} \quad (2.11)$$

Keterangan:

E_i = nilai *Error*

α_i = nilai *alpha* ke-i

D_{ij} = nilai *matriks hessian*

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \quad (2.12)$$

Keterangan:

$\delta\alpha_i$ = nilai *delta alpa*

α_i = nilai *alpha* ke-i

E_i = nilai *error* ke-i

C = nilai konstanta *C*

γ = nilai konstanta *gamma*

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (2.13)$$

Keterangan:

α_i = nilai *alfa* ke-i

$\delta\alpha_i$ = nilai *delta alfa* ke-i

4. Tahapan selanjutnya adalah mendapat nilai *Support Vector (SV)*, $SV = (\alpha_i > \text{ThresholdSV})$. Nilai *ThresholdSV* didapatkan dari beberapa percobaan, biasanya digunakan $\text{ThresholdSV} \geq 0$.

2.5.3 One Against All

Pendekatan *One Against All* adalah salah satu pendekatan yang dapat digunakan dalam menyelesaikan masalah multikelas pada metode SVM. Pada pendekatan *One Against All* dilakukan dekomposisi multikelas sejumlah k SVM biner, dimana jumlah k mewakili jumlah kelas (Hasanah, et al., 2016). Cara kerja pendekatan *One Against All* adalah dengan membuat semua data *training* yang ada pada kelas y_i menjadi bernilai positif, sedangkan data *training* yang bukan berasal dari kelas y_i akan bernilai negatif. Proses ini akan diulang sebanyak jumlah *level* yang digunakan. Selain pendekatan *One Against All* terdapat pendekatan lainnya yaitu *One Against One*.

2.6 K-fold Cross Validation

Pengujian *K-Fold Cross Validation* merupakan suatu proses pengujian untuk menghitung nilai akurasi yang diperoleh dari hasil klasifikasi pada suatu metode. Cara penghitungan pada pengujian ini adalah dengan menghitung rata-rata akurasi setiap data uji yang diacak dan dibagi sebanyak nilai k (Banjarsari, et al., 2015). Dalam pengujian ini banyaknya nilai k menandakan banyaknya subset data yang dibuat, dimana subset ini terdiri dari data yang dibagi secara acak ke dalam setiap subset. Pembagian data latih dan data uji dilakukan dengan cara menggunakan satu subset menjadi data uji dan subset lainnya menjadi data latih. Subset untuk data uji akan dilakukan secara bergantian pada setiap subset k . Setelah itu data uji yang dipakai akan dilihat berapa banyak data uji yang dapat diprediksi benar sesuai dengan kelas pada data asli. Perhitungan untuk mendapatkan nilai akurasi ini adalah dengan membagi jumlah data uji yang sesuai dengan prediksi dengan jumlah seluruh data uji yang digunakan, setelah itu dirata-ratakan untuk sebanyak nilai k . Persamaan untuk perhitungan ini digambarkan pada Persamaan 2.14.

$$\text{Nilai Akurasi} = \frac{\text{jumlah h data akurat}}{\text{jumlah h seluruh h data}} * 100\% \quad (2.14)$$