

## BAB 2 LANDASAN KEPUSTAKAAN

### 2.1 Kajian Pustaka

Pada bagian ini akan menguraikan perihal penelitian yang sudah pernah dilakukan sebelumnya. Penelitian yang dilakukan sebelumnya dapat menggunakan metode selain SVM tetapi dengan objek sama yaitu tingkat risiko pasien dengan penyakit gagal ginjal atau dengan metode SVM tetapi dengan objek berbeda. Bahan pembanding dari penelitian sebelumnya dilihat dari kriteria input, objek, hasil keluaran serta metode yang digunakan.

Penelitian pertama yang dilakukan oleh Rachman mengenai penggolongan tingkat keganasan kanker payudara dengan menggunakan metode *support vector machine* dan analisis regresi logistik ordinal. Variabel input yang digunakan adalah ukuran tumor, *nodus*, *chemotherapy*, *malignant/benign*, letak kanker, usia pasien. Akurasi sistem yang didapat sebesar 98,11%. Sehingga dapat diketahui bahwa metode SVM memiliki ketepatan yang baik untuk klasifikasi daripada regresi logistik ordinal, karena hasil klasifikasi dengan metode regresi ordinal saja mendapatkan akurasi sebesar 56,6% (Rachman, 2012).

Penelitian kedua yang dilakukan oleh Jendi Hardono pada tahun 2013, dengan menggunakan metode *Support Vector Machine* (SVM) untuk klasifikasi pasien penderita penyakit hepatitis. Fitur yang digunakan pada penelitian ini terdiri dari 19 fitur dengan 2 kelas dan jumlah data sebanyak 155 data. Dimana datasetnya diambil dari [www.ics.uci.edu](http://www.ics.uci.edu). Dalam sistem terdapat 4 bagian utama yaitu input data, normalisasi dan kernel data, Training SVM, serta pengujian. Dengan menggunakan metode SVM dengan kernel *linier* maka hasil dari pengujian yang sudah dilakukan diperoleh akurasi sebesar 84,93% (Hardono, 2013).

Penelitian ketiga yang dilakukan oleh munawaroh mengenai klasifikasi tingkat kerusakan jalan pada rel kereta api dengan menggunakan metode *Support Vector Machine* (SVM). Klasifikasi yang dihasilkan berupa rusak berat, rusak sedang dan rusak ringan. Langkah algoritme yang dilakukan adalah normalisasi, menghitung kernel *polynomial d*, perhitungan matriks *hessian*, *sequential training SVM*, menghitung  $f(x)$  serta melakukan klasifikasi dengan strategi *one-against-all*. Sistem yang dibuat dapat melakukan deteksi pada lokasi yang rawan akan kerusakan rel kereta api. Hasil akurasi yang didapat sebesar 91,24% (Munawaroh, 2014).

Penelitian keempat yang dilakukan oleh Shao-Wu Zang mengenai pengukuran protein. Ada beberapa metode yang dilakukan dalam pengukuran protein seperti metode ekstraksi fitur asam dan metode distribusi indeks asam amino. Supaya kedua metode tersebut menghasilkan hasil yang maksimal maka digunakannya fungsi kernel dan Metode SVM. Hal tersebut digunakan untuk menghindari urutan rangkain dua vektor dengan dua protein. Hasil akurasi prediksi sebesar 94% (Zang, Hao & Zhang, 2014). Agar penelitian-penelitian yang dijabarkan di atas mudah dipahami, maka berikut dijelaskan dalam bentuk tabel seperti yang ditunjukkan Tabel 2.1.

Tabel 0.1 Kajian pustaka

No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
1.	(Rachman, 2012)	<p>Obyek: tingkat keganasan <i>breast cancer</i></p> <p>Masukkan:</p> <ul style="list-style-type: none"> <li>- Ukuran tumor</li> <li>- <i>Nodus</i></li> <li>- <i>Cemotherapi</i></li> <li>- <i>Malignant/benign</i></li> <li>- Letak kanker</li> <li>- Usia pasien</li> </ul>	<p>Metode: <i>Support Vector Machine (SVM)</i></p> <p>Proses:</p> <ul style="list-style-type: none"> <li>- Melakukan input data yang diperoleh</li> <li>- Melakukan proses <i>pre-prosessing</i> data</li> <li>- Melakukan proses <i>training SVM</i> dengan menggunakan <i>kernel RBF</i></li> <li>- Membangun pemodelan SVM</li> <li>- Melakukan <i>testing</i> data uji</li> </ul>	<p>Hasil:</p> <p>Hasil dari sistem berupa klasifikasi dalam tingkat stadium I, stadium II, Stadium III. Untuk hasil coba pada penelitian ini didapatkan akurasi sebesar 98,11%</p>
2.	(Hardono, 2013)	<p>Obyek: klasifikasi pasien penderita penyakit hepatitis</p> <p>Masukan:</p> <ul style="list-style-type: none"> <li>- <i>Age</i></li> <li>- <i>Sex</i></li> <li>- <i>Steroid</i></li> <li>- <i>Antiviral</i></li> </ul>	<p>Metode: <i>Support Vector Machine (SVM)</i></p> <p>Proses:</p> <ul style="list-style-type: none"> <li>- Melakukan normalisasi pada setiap atribut</li> </ul>	<p>Hasil:</p> <p>Hasil coba pada penelitian tersebut diperoleh akurasi yang tinggi sebesar 84,93%</p>

No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
		<ul style="list-style-type: none"> <li>- <i>Fatigue</i></li> <li>- <i>Malaise Anorexia</i></li> <li>- <i>Liver Big</i></li> <li>- <i>Liver Firm</i></li> <li>- <i>Spleen</i></li> <li>- <i>Palpable</i></li> <li>- <i>Spiders</i></li> <li>- <i>Ascites</i></li> <li>- <i>Varices</i></li> <li>- <i>Bilirubin</i></li> <li>- <i>Alg</i></li> <li>- <i>Phosphate</i></li> <li>- <i>SGot</i></li> <li>- <i>Albumin</i></li> <li>- <i>Protine</i></li> <li>- <i>Hystologi</i></li> </ul>	<ul style="list-style-type: none"> <li>- Melakukan proses <i>training SVM</i> dengan menggunakan <i>kernel linier</i></li> <li>- Mencari nilai <i>w, b, slack variable</i> menggunakan <i>quadratic programming</i></li> <li>- Melakukan pengujian <i>SVM</i> dengan menghitung <i>score</i> dan nilai dari klasifikasi</li> </ul>	
3.	(Maghfirah, 2014)	<p>Obyek: klasifikasi tingkat risiko penyakit gagal ginjal</p> <p>Masukkan:</p> <ul style="list-style-type: none"> <li>- Umur</li> <li>- Urea</li> <li>- BUN</li> <li>- Kreatinin</li> </ul>	<p>Metode: <i>Fuzzy K-Nearest Neighbor</i></p> <p>Proses:</p> <ul style="list-style-type: none"> <li>- Menentukan fungsi derajat keanggotaan menggunakan <i>fuzzy</i></li> </ul>	<p>Hasil:</p> <p>Hasil yang didapat dari penelitian tersebut adalah berupa nilai <i>f-measure</i> dari prediksi untuk setiap data kedalam data penyakit gagal ginjal. Dimana nilai <i>f-measure</i></p>

No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
		<ul style="list-style-type: none"> <li>- <i>Uric acid</i></li> <li>- Trigliserida</li> </ul>	<ul style="list-style-type: none"> <li>- dari setiap parameter</li> <li>- Melakukan normalisasi dengan <i>fuzzy</i></li> <li>- Mencari jarak <i>Euclidean distance</i></li> <li>- Melakukan pengurutan dari jarak terkecil hingga terbesar</li> <li>- Memasukkan nilai <i>k</i></li> <li>- Melakukan <i>sorting</i> data</li> <li>- Menghitung nilai <i>F-Measure</i></li> </ul>	<p>memiliki prediksi tertinggi saat nilai <math>k=10</math> dengan hasil nilainya 0,8 dari sebaran data latih sebanyak 60, 90 dan 120.</p>
4.	(Hasanah, 2016)	<p>Obyek: Klasifikasi kondisi detak jantung</p> <p>Masukkan :</p> <ul style="list-style-type: none"> <li>- Data sinyal MIT-BIH <i>Arrhythmia database</i></li> </ul>	<p>Metode: <i>Support Vector Machine</i> (SVM)</p> <p>Proses:</p> <ul style="list-style-type: none"> <li>- Melakukan input <i>dataset</i></li> <li>- Melakukan perhitungan data</li> </ul>	<p>Hasil:</p> <p>Hasil dari percobaan dalam penelitian ini adalah berupa hasil akurasi pada klasifikasi kondisi detak jantung yang diambil dari hasil rekam elektrokardiografi. Hasil akurasi didapat pada <i>scenario</i> pengujian yang ke-4</p>

No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
			<p>dengan <i>kernel polynomial</i></p> <ul style="list-style-type: none"> <li>- Melakukan perhitungan data latih dengan SVM menggunakan metode <i>sequential training SVM</i></li> <li>- Menentukan kelas data pada level menggunakan strategi <i>One-Againts-All</i></li> <li>- Melakukan prediksi dari data uji</li> </ul>	<p>dengan hasil sebesar 87,5% dan rasio data nya 90%:10%.</p>
5.	(Munawaroh, 2014)	<p>Obyek: Klasifikasi tingkat kerusakan jalan pada rel kereta api</p> <p>Masukkan:</p> <ul style="list-style-type: none"> <li>- Jenis rel</li> <li>- Sambungan rel putus</li> <li>- Rel <i>defect</i> (cacat)</li> <li>- Insulator rusak</li> <li>- Railpad rusak</li> </ul>	<p>Metode: <i>Support Vector Machine</i> (SVM)</p> <p>Proses:</p> <ul style="list-style-type: none"> <li>- Melakukan inialisasi dan normalisasi</li> <li>- Melakukan perhitungan data</li> </ul>	<p>Hasil : Klasifikasi yang dihasilkan rusak berat, rusak sedang, dan rusak ringan, dengan akurasi sebanyak 91,24%</p>

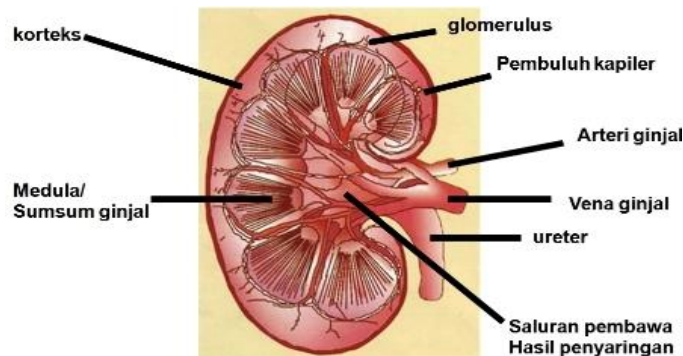
No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
		<ul style="list-style-type: none"> <li>- Plat sambung</li> <li>- Baud sambung</li> <li>- Penambat</li> <li>- Ballast mati</li> </ul>	<ul style="list-style-type: none"> <li>dengan kernel <i>polynomial d</i></li> <li>- Melakukan perhiungan <i>matriks hessian</i></li> <li>- Perhitungan <i>sequential training SVM</i></li> <li>- Mengitung nilai <math>E_i</math>, <math>\delta\alpha_i</math>, <math>\alpha_i</math> baru</li> <li>- Melakukan <i>testing SVM</i></li> <li>- Mencari nilai bias dan <math>f(x)</math></li> <li>- Melakukan proses klasifikasi dengan strategi <i>One-Againts-All</i></li> </ul>	
6.	(Zang, Hao & Zhang, 2014)	Obyek: Pengukuran Protein	Metode: Fungsi <i>Kernel</i> dan <i>Support Vector Machine</i>	Hasil : Ekstraksi dengan menggunakan fungsi <i>kernel</i> dan metode <i>Support Vector Machine</i> menghasilkan hasil ekstraksi

No.	Peneliti	Obyek(Masukkan)	Metode(Proses)	Hasil(Keluaran)
				yang maksimal dan metode AAID memiliki akurasi sebesar 94%
7.	(Abbaszadeh,Hez arkhani & Mohammadi, 2015)	Obyek: pemisahan zona perubahan sebagai parameter kontrol pada tembaga  Masukkan : <ul style="list-style-type: none"> <li>- Gamma (<math>\gamma</math>)</li> <li>- Complexity C</li> </ul>	Metode: <i>Support Vector Machine</i>	Hasil :  Metode SVM dapat dengan efektif untuk memisahkan perubahan zona. Dan dapat memberikan teknik klasifikasi yang ekonomis, efisien, dan data yang didapat valid
8.	(Darsyah, 2014)	Obyek : Klasifikasi Tuberkulosis  Masukkan : <ul style="list-style-type: none"> <li>- Umur</li> <li>- Berat badan</li> <li>- Tinggi badan</li> </ul>	Metode : <i>Support Vector Machine</i>	Hasil :  Hasil dengan klasifikasi menggunakan metode SVM adalah metode SVM mampu mengukur ketepatan dalam klasifikasi dengan akurasi yang lebih tinggi. Hasil akurasi yang diperoleh dengan fungsi kernel RBF sebesar 98%

## 2.2 Ginjal

### 2.2.1 Definisi Ginjal

Salah satu organ tubuh manusia yang paling penting untuk dirawat adalah ginjal. Organ ini berfungsi mempertahankan kestabilan biologis pada tubuh, dimana peran dari ginjal meliputi pengaturan cairan di dalam tubuh, keseimbangan elektrolit, mengeluarkan hasil metabolit ekskresi obat dari dalam tubuh manusia (Lucida, Trisnawati & Suardi, 2011). Manusia memiliki ginjal sebanyak dua buah yang terletak sedikit dibawah tulang rusuk bagian belakang. Ginjal manusia memiliki ukuran panjang sekitar 7cm dan tebal 3cm serta dibungkus dengan kapsul yang terbuka kearah bawah. Terdapat jaringan lemak diantara ginjal dan kapsul yang memiliki fungsi untuk melindungi ginjal terhadap goncangan. Ginjal mengeluarkan semacam zat untuk memperbaiki aliran darah yang mengalami gangguan. Biasanya sekitar 20% darah manusia akan mengalir melewati ginjal agar dibersihkan selama waktu 1 menit. Darah tersebut masuk melalui pembuluh nadi ginjal menuju ke jaringan ginjal yang bercabang-cabang sampai berubah menjadi kapiler dan darah akan mencapai gromelurus. Glomelurus berbentuk seperti gelas yang digunakan untuk minum anggur dan pembuluh kapiler akan mengisi bagian gelas itu. Salah satu ciri ginjal mengalami kerusakan adalah ginjal mengalami kesulitan dalam mengeluarkan cairan tubuh dan zat-zat sisa pembakaran yang beracun. Sehingga tubuh penderita yang terletak lebih renah mengalami pembengkakan dan penderita akan mengalami peningkatan tekanan darah (Aria & Suarna, 2009). Untuk struktur ginjal pada manusia ditunjukkan pada Gambar 0.1.



**Gambar 0.1 Struktur ginjal pada manusia**

Sumber: (Sari, 2015)

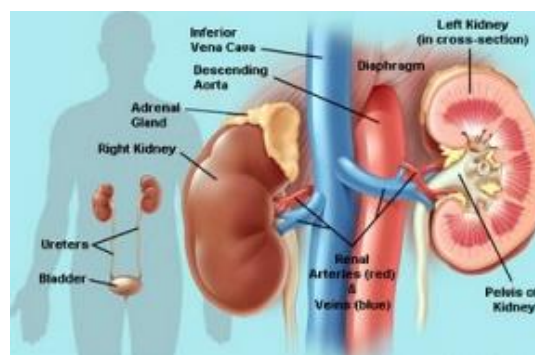
### 2.2.2 Gagal Ginjal

Gagal ginjal merupakan suatu tanda bahwa organ ginjal dalam melakukan penyaringan pembuangan elektrolit dalam tubuh mengalami penurunan serta tidak lagi berfungsi dengan baik dalam menjaga zat kimia dan keseimbangan cairan kalium juga sodium yang ada di dalam darah saat produksi urin. Penyakit tersebut bisa berkembang ke kondisi yang lebih buruk lagi dan dapat membuat ginjal tidak lagi bekerja sebagaimana mestinya (Warianto, 2011). Gagal ginjal terdiri dari dua bagian meliputi gagal ginjal akut dan gagal ginjal kronik. Bagian



pertama, yaitu gagal ginjal akut yang mana dapat dilihat dari adanya kesulitan dalam buang air kecil atau dapat dikatakan bahwa volume *urine* semakin sedikit dan mengalami rasa sakit ketika buang air kecil, serta air dapat mencapai paru-paru sehingga membuat penderita susah tidur. Gagal ginjal akut ini juga ditandai dengan meningkatnya kreatinin dan pembengkakan pada kaki karena penuh air (Harahap, 2014).

Penyakit ginjal kronik (PGK) merupakan terjadinya suatu kondisi penurunan fungsi ginjal yang bersifat kronik, progresif dan biasanya berlangsung lambat dalam beberapa tahun. Pada kondisi ini ginjal akan kehilangan kemampuannya dalam mempertahankan komposisi dan volume cairan tubuh dalam kondisi asupan makanan normal (Haryanti & Nisa, 2015). Gagal ginjal kronik dapat terjadi secara perlahan dengan hitungan bulan ataupun tahun, serta yang paling berbahaya sifatnya tidak dapat disembuhkan (Nurani & Mariyanti, 2013). Terdapat empat tindakan yang akan mengurangi perkembangan penyakit ginjal kronik (PGK) meliputi mengontrol tekanan darah dengan patokan tekanan darah < 140/90 mm/Hg, penggunaan obat-obatan seperti *angiotensin converting enzyme* inhibitor serta *angiotensin receptor blocker* albumin dan hipertensi, mengontrol kadar diabetes, melakukan kontrol asidosis metabolik atau kadar keasaman dalam darah (Vassalotti & et.al, 2015). Penderita gagal ginjal yang pada stadium akhir dalam mempertahankan kelangsungan hidupnya maka diperlukan proses terapi yaitu hemodialysis (HD) dan transplantasi ginjal. Tanda bahwa klirens kreatinin rendah muncul sebagai tanda bahwa penderita gagal ginjal berada pada stadium akhir. Jika penderita memiliki klirens kreatinin dengan nilai 15 ml/menit maka disarankan agar menjalani terapi pengganti seperti dialisis. Terapi dialisis dilakukan bertujuan untuk menurunkan kadar ureum, kreatinin dan zat toksin lainnya didalam darah. Penyakit gagal ginjal ini tidak pilih-pilih pasien, penyakit tersebut dapat menyerang siapa saja apalagi pada seseorang yang sedang menderita penyakit serius sehingga langsung dapat berdampak ke ginjal. Biasanya seseorang yang berusia dewasa dan lanjut usia rentan menderita penyakit gagal ginjal (Warianto, 2011). Salah satu bagian dari ginjal manusia yang tidak berfungsi, sebagai tanda bahwa terjadinya gagal ginjal ditunjukkan pada Gambar 0.2.



**Gambar 0.2** Salah satu bagian ginjal manusia yang tidak berfungsi, sebagai tanda bahwa terjadinya gagal ginjal

Sumber: (Sari, 2015)

### 2.2.3 Faktor yang Memengaruhi Risiko Penyakit Gagal Ginjal

#### 1. Jenis Kelamin

Jenis kelamin merupakan konsep analisis yang digunakan untuk melakukan identifikasi perbedaan dari laki-laki dan perempuan dengan dilihat melalui aspek sudut pandang non-biologis yaitu aspek sosial, psikologis maupun budaya (Normadewi, 2012). Penyakit gagal ginjal sering dialami oleh laki-laki daripada perempuan, karena laki-laki memiliki risiko 2 kali lebih besar terkena penyakit gagal ginjal daripada perempuan. Hal tersebut dapat dijelaskan bahwa laki-laki kurang memperhatikan kesehatan dan menjaga pola hidupnya.

#### 2. Umur

Umur merupakan waktu ketika seseorang hidup atau waktu dari seseorang tersebut dilahirkan. Meningkatnya umur berarti rentan dalam penurunan kualitas hidup. Pasalnya dengan meningkatnya hidup seseorang akan rentan dengan berbagai penyakit yang akan dialami. Seseorang yang memiliki umur dewasa dan tua sering atau lebih rentan mengalami suatu penyakit berbahaya yaitu gagal ginjal (Butar-butar & Siregar, 2009). Variabel Umur dibagi menjadi 3 kategori, yaitu muda < 35 tahun, dewasa 35-55 tahun, tua > 55 tahun (Kusumadewi, 2004).

#### 3. Kadar Urea

Urea merupakan hasil sisa metabolisme tubuh yang dikeluarkan oleh ginjal serta berasal dari diet protein. Batas normal dari kadar urea yaitu 10-50 mg/dl. Jika kadar urea melebihi batas normal maka fungsi ginjal dalam kondisi bahaya (Suryawan, Arjani & Sudarmanto, 2016).

#### 4. Kadar BUN

*Blood Urea Nitrogen* (BUN) merupakan parameter yang digunakan untuk menilai fungsi ginjal serta mengukur jumlah nitrogen urea yang hadir dalam darah. Karena BUN langsung berhubungan dengan ekskresi fungsi ginjal maka jika ginjal mengalami permasalahan maka kadar BUN dalam darah akan meningkat. Batas normal untuk kadar BUN, yaitu 6-20 mg/dl (Kamal, 2014).

#### 5. Kadar kreatinin

Kreatinin merupakan hasil sisa dari penguraian fosfat kreatinin yang terjadi pada otot. Kreatinin ini disebut juga zat racun dalam darah, serta berada pada seseorang yang ginjalnya sudah rusak atau tidak berfungsi dengan normal. Batas normal kadar kreatinin dalam darah yaitu untuk perempuan 0,6-1,1 mg/dl sedangkan untuk laki-laki yaitu 0,7-1,3 mg/dl (Marzuki, Sennang & Fematubun, 2014).

#### 6. Kadar Uric Acid

Uric acid adalah suatu hasil akhir metabolisme purin yang berasal dari metabolisme dalam tubuh/faktor genetik dan dari luar tubuh/faktor sumber makanan yang dikonsumsi. Kadar uric acid yang tinggi pada darah, dapat

merusak organ tubuh salah satunya adalah ginjal, karena hal itu akan menyebabkan penyumbatan saringan dalam ginjal serta akan mengakibatkan gagal ginjal (Lina & Setiyono, 2014). Batas normal uric acid pada laki-laki 3,4-7,0 mg/dl dan perempuan 2,4-5,7 mg/dl (Silviana, Bintanah & Isworo, 2015).

#### 7. Kadar Trigliserida

Kadar trigliserida merupakan jenis lemak yang berada dalam darah dan terdapat pada organ tubuh lainnya. Pembentukan trigliserida dari lemak dan gliserol hasil uraian makanan yang telah dikonsumsi masuk ke tubuh serta menuju ke hati. Meningkatnya kadar trigliserida akan membahayakan kesehatan karena zat dalam darah akan tertimbun pada dinding pembuluh darah (Arifnaldi, 2014). Kadar trigliserida darah: normal= jika kurang dari 200 mg/dl, garis batas tinggi= 200-250 mg/dl, tinggi= lebih dari 250 mg/dl.

### 2.3 Data Mining

Ada beberapa pendapat tentang pengertian dari *data mining*, berikut penjelasan mengenai pendapat *data mining* secara umum:

1. *Data mining* atau istilah yang lainnya adalah *knowledge discovery in database* (KDD) merupakan suatu aktivitas dalam pengumpulan, pemakaian data historis yang digunakan untuk menemukan keteraturan pola, dimana pola dan hubungan berada pada satu set yang memiliki ukuran data yang besar. (Santosa, 2007)
2. *Data mining* dapat diartikan sebagai suatu proses untuk menemukan hubungan, trend dan pola yang mempunyai makna. Proses tersebut dilakukan dengan menggunakan teknik pengenalan pola contohnya seperti teknik matematika dan teknik statistika. Cara pertama dengan proses penyaringan data yang memiliki ukuran yang besar, dimana datanya disimpan di penyimpanan yang sudah ditentukan. (Kusnawi, 2007)
3. *Data mining* adalah suatu proses dalam menentukan pengetahuan dari sejumlah data yang besar kemudian disimpan dalam *database, warehouse*, atau tempat penyimpanan informasi yang lain. (Han & Kamber, 2006)

Jadi secara garis besar *data mining* adalah suatu istilah yang berhubungan dengan *knowledge discovery* (penemuan pengetahuan) dan *pattern recognition* (pengenalan pola). Hal tersebut dikaitkan karena *data mining* memiliki tujuan untuk mendapatkan pengetahuan yang masih tersembunyi atau belum diketahui dari dalam kumpulan data. Pengetahuan tersebut juga dapat digunakan untuk menggali data-data yang masih berbentuk pola. Sehingga *data mining* adalah suatu proses untuk menemukan dan menggali sebuah pengetahuan dari kumpulan data maupun informasi. (Susanto & Suryadi, 2010)

### 2.4 Klasifikasi

Klasifikasi merupakan sebuah proses dalam pengelompokan dari sejumlah data pada kelas tertentu. Untuk penentuan kelasnya diberikan sesuai dengan sifat

dan pola yang dimiliki oleh data tersebut. Terdapat dua pekerjaan utama dalam proses klasifikasi, yaitu pembangunan sebuah model yang akan dijadikan rancangan awal untuk disimpan sebagai memori, kemudian model itu akan digunakan untuk melakukan proses klasifikasi/prediksi/pengenalan pada suatu obyek data supaya diketahui berada dikelas mana, dengan model yang sudah ditentukan sebelumnya (Susanto & Suryadi, 2010). Menurut Kusnawi (2007), klasifikasi merupakan suatu proses untuk menentukan objek kedalam satu kelas yang sebelumnya telah didefinisikan. Klasifikasi juga dapat digunakan untuk menentukan kelas pada objek yang belum diketahui labelnya. Teknik yang digunakan dalam melakukan klasifikasi data baru, yaitu melalui cara memanipulasi data yang sudah diklasifikasi terlebih dahulu kemudian hasilnya diberikan sejumlah aturan. Sejumlah aturan tersebut diterapkan pada data-data baru untuk dilakukan proses klasifikasi. Salah satu metode klasifikasi yang memiliki kinerja sangat baik pada data berdimensi tinggi serta dapat dengan baik memetakan data asli dari dimensi asal menjadi dimensi lainnya yang lebih tinggi dengan menerapkan teknik kernel adalah metode *Support Vector Machine* (SVM) (Budiarti, 2017).

## .5 Analisis Korelasi

Korelasi didefinisikan sebagai suatu teknik analisis yang ada dalam dunia statistika dengan tujuan untuk mencari hubungan antara dua variabel atau lebih. Hubungan tersebut dapat terjadi ketika adanya hubungan sebab akibat atau karena kebetulan saja. Dikatakan dua variabel atau lebih berkorelasi dilihat dari perubahan pada variabel satu akan diikuti perubahan variabel yang lain secara teratur dengan arah yang sama (korelasi positif) atau arah yang berlawanan (korelasi negative). Pada penelitian ini menggunakan jenis korelasi sederhana, yaitu dengan mengukur kekuatan atau keterikatan hubungan antar dua variabel. Kekuatan antar dua variabel didefinisikan erat, lemah atau tidak erat (tidak ada hubungannya sama sekali) (Steel & Torrie, 1989). Dalam korelasi terdapat koefisien korelasi, yang mana koefisien tersebut menggambarkan tingkat keeratan antara hubungan linier dua variabel. Koefisien tersebut dinotasikan sebagai  $r$  serta memiliki nilai berkisar antara -1 sampai 1 ( $-1 \leq r \leq 1$ ) yang dapat dijelaskan sebagai berikut (Pramoedyo, 2013):

- a. Apabila korelasi ( $r$ ) = -1, derajat keterkaitan hubungan antara dua variabel sangat kuat dan mempunyai hubungan negatif (berlawanan arah).
- b. Apabila korelasi ( $r$ ) = 1, derajat keterkaitan hubungan antara dua variabel sangat kuat dan mempunyai hubungan positif (searah).
- c. Apabila korelasi ( $r$ ) = 0, hubungan antara dua variabel tidak memiliki hubungan sama sekali (hubungan  $x$  dan  $y$  lemah).

Berikut adalah rumus dari pencarian nilai koefisien relasi yang ditunjukkan pada persamaan 2.1 (Pramoedyo, 2013):

$$r = \frac{n \sum_{i=1}^n x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}} \quad (2.1)$$

Keterangan:

$r$  = koefisien korelasi

$n$  = jumlah sampel

$x$  = nilai variabel  $x$

$y$  = nilai variabel  $y$

## 1.6 Normalisasi

Normalisasi digunakan supaya nilai antar fitur berada pada, interval yang sama. *Min-Max Normalization* adalah metode normalisasi yang sering digunakan untuk mengatasi permasalahan nilai antar fitur yang memiliki jarak terlampau jauh. Rumus dari *Min-Max Normalization* ditunjukkan pada Persamaan 2.2 (Larose, 2005).

$$X^* = \left( \frac{X - \min(X)}{\max(X) - \min(X)} \right) \quad (2.2)$$

Keterangan:

$X^*$  : hasil normalisasi data

$X$  : data yang belum dinormalisasi

$\min(X)$  : nilai minimum dari semua data

$\max(X)$  : nilai maksimum dari semua data

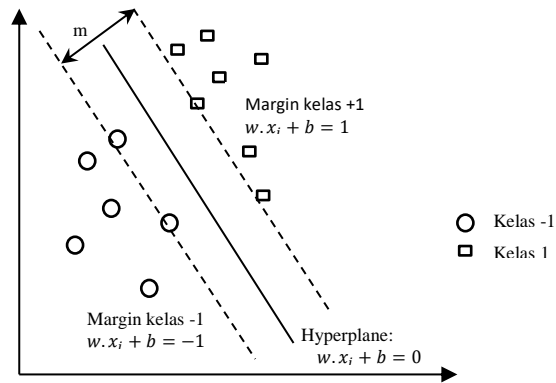
Berikut tahapan proses normalisasi untuk data latih dan data uji:

1. Langkah awal masukkan berupa data latih dan data uji.
2. Langkah kedua melakukan pencarian nilai maksimum serta nilai minimum dari setiap fitur yang terdapat pada data uji dan data latih.
3. Kemudian menggunakan *Min-Max Normalization* untuk melakukan proses perhitungan.

Selanjutnya hasil normalisasi dari masing-masing fitur didapat dari perhitungan *Min-Max Normalization* tersebut.

## 2.7 Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah salah satu metode klasifikasi yang menerapkan teori pembelajaran statistika dan dapat memberikan hasil klasifikasi yang baik. Metode SVM tergolong metode klasifikasi yang bekerja dengan baik pada kumpulan data yang memiliki dimensi tinggi, serta dapat dengan baik memetakan data asli dari dimensi asal menjadi dimensi lainnya yang lebih tinggi dengan menerapkan teknik kernel. Dan juga memiliki konsep memisahkan dua buah kelas dan berusaha untuk menemukan *hyperplane* terbaik (Prasetyo, 2012) langkah tersebut diilustrasikan pada Gambar 0.3.



**Gambar 0.3 Hyperplane SVM**

Sumber: (Prasetyo, 2012)

Untuk menemukan *hyperplane* terbaik antara dua kelas adalah dengan cara mengukur margin dari *hyperplane* tersebut dan mencari titik maksimalnya. Margin didefinisikan sebagai jarak antara *hyperplane* dengan data yang dekat dari masing-masing kelas. Data yang dekat itu biasanya disebut sebagai *support vector* (Prasetyo, 2012). Margin yang optimal dihitung dengan rumus  $1/||w||$  (Nugroho, 2007).

### 2.7.1 SVM Non Linier

Ada beberapa permasalahan yang tidak dapat diklasifikasi menggunakan metode SVM *linier*. Sehingga dapat diklasifikasi dengan metode SVM *non linier*, yang mana menerapkan fungsi kernel untuk menyelesaikan permasalahan SVM *non linier* tersebut. Macam-macam fungsi kernel dalam SVM antara lain (Vijayakumar, 1999):

1. *Kernel* linier

$$K(x, y) = x.y \tag{2.3}$$

2. *Kernel* Polinomial of degree

$$K(x, y) = (x.y)^d \tag{2.4}$$

3. *Kernel* Polinomial of degree 2

$$K(x, y) = (x.y + c)^d \tag{2.5}$$

4. *Kernel* Gaussian RBF

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{2.\sigma^2}\right) \tag{2.6}$$

5. *Kernel* Sigmoid (Tangen Hiperbolik)

$$K(x, y) = \tanh(\sigma(x.y) + c) \tag{2.7}$$

6. *Kernel* Invers Multi Kuadratik

$$K(x, y) = \frac{1}{\sqrt{\|x - y\|^2 + c^2}} \quad (2.8)$$

### 7. Kernel Additive

$$K(x, y) = \sum_{i=1}^n K_1(x_i, y_i) \quad (2.9)$$

Keterangan:

$K(x, y)$	: kernel
$x$	: data $x$
$y$	: data $y$
$d$	: jarak <i>hyperplane</i>
$c$	: complexity
$\sigma$	: sigma

### 2.7.2 Sequential Training

Metode SVM memiliki berbagai macam proses training meliputi *Quadratic Programming (QP)*, *Sequential Training*, dan *Sequential Minimal Optimization (SMO)*. Penyelesaian masalah dengan menggunakan proses *training* QP memiliki algoritme yang cukup kompleks sehingga membutuhkan proses *training* dengan waktu yang lama. Proses *training* SMO merupakan hasil pengembangan dari proses *training* QP, dimana untuk SMO digunakan dalam menyelesaikan permasalahan dengan optimasi yang kecil di setiap tahapannya tetapi algoritme yang digunakan kompleks. Sedangkan untuk proses *training Sequential Training* waktu yang dibutuhkan untuk menyelesaikan permasalahan lebih cepat dikarenakan memiliki Algoritme yang sederhana. Berikut penjelasan dari Algoritme *sequential training* (Vijayakumar, 1999):

1. Melakukan inialisasi  $\alpha_i$  dan parameter yang lain, yaitu  $\lambda, \gamma, C$ , dan  $\varepsilon$ .

Keterangan:

$\alpha_i$	: lagrange multiplier, berguna untuk mencari <i>support vector</i> .
$\gamma$	: <i>gamma</i> , berguna dalam melakukan <i>controlling</i> kecepatan <i>training</i> .
$C$	: complexity $C$ , parameter ini digunakan dalam proses <i>training</i> yang nantinya memberikan batas nilai alfa.
$\varepsilon$	: <i>epsilon</i> , digunakan dalam mengukur <i>error</i> pada klasifikasi

2. Melakukan perhitungan matriks *hessian*

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (2.10)$$

Untuk  $i, j = 1, \dots, n$

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*

Setelah selesai melakukan proses penghitungan matriks *hessian*, maka selanjutnya menggunakan persamaan 2.11, persamaan 2.12 dan persamaan 2.13 untuk menghitung proses  $E_i$ ,  $\delta\alpha_i$  dan  $\alpha_i$  pada data yang dipakai mulai dari data ke- $i$  sampai ke- $j$ .

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

Keterangan:

- $\alpha_i$  : alfa ke- $j$
- $D_{ij}$  : matrik *hessian*
- $E_i$  : *error rate*

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

Keterangan:

- $\alpha_i$  : alfa ke- $i$
- $\gamma$  : *gamma*
- $E_i$  : *error rate*
- $C$  : nilai kompleksitas

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

- $\alpha_i$  : alfa ke- $i$
- $\delta\alpha_i$  : *delta alfa* ke- $i$

3. Pada tahap selanjutnya proses dari  $E_i$ ,  $\delta\alpha_i$  dan  $\alpha_i$  dilakukan berulang-ulang sampai  $\max(|\delta\alpha_i|) < \varepsilon$  atau selama kondisi iterasi maksimum tercapai.
4. Kemudian yang terakhir didapatkan nilai *support vector* (SV),  $SV = (\alpha_i > \text{ThresholdSV})$ , dengan memberikan nilai  $\text{ThresholdSV} = 0$ .

## 2.8 Testing SVM

Tahap awal yang dilakukan adalah mencari nilai  $f(x)$ , dengan rumus dalam Persamaan 2.14.



$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (2.14)$$

Keterangan:

- $\alpha_i$  : alfa ke- $i$
- $y_i$  : kelas data latih ke- $i$
- $m$  : jumlah data
- $K(x_i, x)$  : fungsi *kernel* yang dipakai

Dimana  $x$  : data uji

$x_i$  : data latih ke- $i$

$b$  : *bias*

Nilai  $b$ (*bias*) dicari menggunakan rumus dalam Persamaan 2.15

$$b = -\frac{1}{2} \left[ \sum_{i=1}^m \alpha_i y_i K(x_i, x^+) + \sum_{i=1}^m \alpha_i y_i K(x_i, x^-) \right] \quad (2.15)$$

Keterangan:

- $\alpha_i$  : alfa ke- $i$
- $y_i$  : kelas data latih ke- $i$
- $m$  : jumlah data
- $K(x_i, x^+), K(x_i, x^-)$  : fungsi *kernel* yang dipakai, dimana:
  - $x^+$  = nilai alfa yang paling besar pada kelas positif yang terdapat pada data latih
  - $x^-$  = nilai alfa yang paling kecil pada kelas positif yang terdapat pada data latih
  - $x_i$  = data latih ke- $i$
- $b$  : *bias*

## 2.9 Multiclass SVM

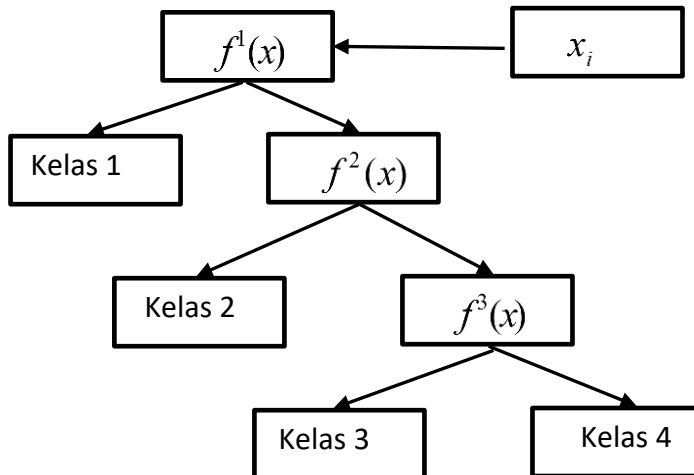
SVM merupakan metode yang biasanya digunakan dalam mengatasi permasalahan yang memiliki dua kelas. Dalam pengambilan keputusan untuk permasalahan *multi-class problem* digunakan dua metode yaitu klasifikasi *multi-class* dimana metode ini melakukan pendekatan dengan menggabungkan dan mengkombinasikan *binary classifier* terhadap semua data pada permasalahan optimasi yang terdiri dari beberapa kelas (Sembiring, 2007). Pada umumnya yang digunakan dalam menyelesaikan permasalahan tersebut adalah mengkombinasikan beberapa dari *binary classifier*. Beberapa jenis metode dalam *multiclass SVM* antara lain *Directed Acyclic Graph Support Vector Machine* (DAGSVM), *one-against-one*, *one-against-all*, dan *Binary Decision Tree*(BDT).

## 2.10 Metode *One-Againts-All*

*Multiclass SVM* yang dikembangkan sesuai dengan sejumlah  $k$ , dimana  $k$  tersebut merupakan jumlah dari kelas, yaitu pendekatan dengan *one-againts-all* (Sari, 2016). Seperti contoh jika terdapat permasalahan tentang klasifikasi dengan jumlah kelasnya empat buah, kemudian untuk pelatihannya digunakan tiga buah SVM yang ditunjukkan pada Tabel 0.2 dan digunakannya metode tersebut dalam klasifikasi data baru ditunjukkan pada Gambar 0.4.

**Tabel 0.2 Metode *One-Againts-All* dengan 4 kelas**

$y_1 = 1$	$y_1 = -1$	Hipotesis
Kelas 1	Bukan kelas 1	$f^1(x) = (w^1)x + b^1$
Kelas 2	Bukan kelas 2	$f^2(x) = (w^2)x + b^2$
Kelas 3	Kelas 4	$f^3(x) = (w^3)x + b^3$



**Gambar 0.4 Contoh klasifikasi metode *One-Againts-All* dengan 4 kelas**

## 2.11 Akurasi Hasil Pengujian

Akurasi dari hasil pengujian diperoleh dari jumlah data uji yang hasil klasifikasinya memiliki kesamaan dengan kelas sebenarnya. Tingkat akurasi dihitung menggunakan rumus pada persamaan 2.16.

$$Akurasi(\%) = \frac{\Sigma \text{data uji sesuai}}{\Sigma \text{total data uji}} \times 100\% \quad (2.16)$$

$\Sigma \text{ data uji sesuai}$ : jumlah data uji hasil klasifikasi yang memiliki kesamaan dengan *actual classnya*.

$\Sigma \text{ total data uji}$  : jumlah data uji keseluruhan.