

BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. TINJAUAN PUSTAKA

Khoirul (2013) pengenalan suara menggunakan metode *Linier Predictive Coding* (LPC) bertujuan untuk mendapatkan beberapa koefisien LPC yang merupakan ciri *featur* ciri suara manusia. Kemudian koefisien LPC diproses dengan menggunakan *Fast Fourier Transform* (FFT) untuk mendapatkan sinyal pada domain frekuensi. Hal ini bertujuan agar perbedaan antar pola kata yang satu dengan yang lain terlihat lebih jelas sehingga ekstraksi parameter memberikan hasil yang lebih baik.

Zulham Effendi, dkk (2015) Pengenalan Suara Menggunakan Metode MFCC (*Mel Frequency Cepstrum Coefficients*) dan DTW (*Dynamic Time Warping*) untuk Sistem Penguncian Pintu penelitian ini merancang sebuah sistem keamanan penguncian pintu menggunakan suara. Sistem ini menggunakan metode MFCC (*Mel Frequency Cepstrum Coefficients*) untuk mengekstraksi ciri sinyal suara dan metode DTW (*Dynamic Time Warping*) untuk mencocokkan pola suara. Mikrofon *portable* digunakan untuk meng-input-kan suara dan pengolahan dilakukan dengan menggunakan komputer. Pertama, ciri suara diekstraksi, kemudian dilakukan pencocokan pola suara antara sinyal uji dengan sinyal referensi. Jika skor *normalized distance* yang dihasilkan < 80 , maka akan dikirim data “1” ke mikrokontroler dan pintu akan terbuka. Selain itu, data “0” akan dikirim ke mikrokontroler dan pintu tetap terkunci. Pengujian dilakukan oleh 4 *trainer* yang mengucapkan kata uji “*pintu buka*” dan kata uji acak. Tingkat keberhasilan untuk kata uji „*pintu buka*” yaitu sebesar 81,25 %. Sedangkan untuk kata uji acak, memiliki tingkat keberhasilan sebesar 57,5 %.

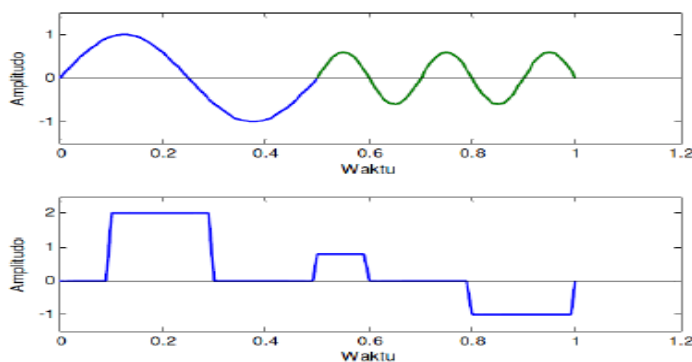
Prativi Nugraheni Hanggarsar, dkk (2012) simulasi sistem pengacakan sinyal suara secara realtime berbasis *Fast Fourier Transform* (FFT) telah melakukan simulasi data pengacakan sinyal suara berbasis FFT. Sistem ini melakukan perekaman sinyal suara secara *realtime* sebagai sinyal input melalui *microphone* yang tersedia pada perangkat komputer. Dalam proses enkripsi, sinyal informasi yang berupa sinyal analog (domain waktu) dikonversikan ke dalam domain frekuensi menggunakan algoritma *Fast Fourier Transform* (FFT) 16 titik dan 32 titik. Sistem pembicaraan aman dengan menggunakan FFT 32 titik memiliki waktu proses yang lebih lama dibandingkan FFT 16 titik. Untuk sistem yang menggunakan FFT 16 titik membutuhkan waktu rata-rata 0,18564 detik sedangkan pada FFT 32

titik membutuhkan waktu rata-rata 2,70816 detik dengan selisih waktu 2.55252 detik.

2.2. DASAR TEORI

2.2.1. Pengertian Suara

Suara adalah sesuatu yang dihasilkan oleh getaran yang berasal dari benda bergerak, sesuatu yang menghasilkan bunyi, dan tekanan gelombang udara, maka memiliki nilai kontinu terhadap waktu (analog). Suara biasanya merambat melalui udara. Tidak bisa merambat melalui ruang hampa. Gelombang suara bervariasi dalam tingkatan tekanan suara (amplitudo) dan dalam frekuensi. Jumlah waktu yang diperlukan untuk terjadinya suatu getaran atau gelombang dinamakan perioda (T). Sedangkan jumlah gelombang yang terjadi setiap detik dinamakan frekuensi (f) dengan satuan m/dt (Hz) pada contoh gambar 2.1.



Gambar 2.1 Sinyal suara (Reonaldo Yohanes Sipasulta dkk, 2014).

Sebuah sinyal yang rumit sebagai sebuah hasil dari beberapa transformasi yang terjadi pada beberapa level yang berbeda dari semantik, linguistik, artikulasi (pengucapan) dan akustik. Perbedaan dalam transformasi ini tampak sebagai perbedaan anatomic yang melekat dalam *vocal tract* dan kebiasaan pengucapan yang dipelajari dari individu yang berbeda (Ronando dkk, 2012).

2.2.2. Pola Suara

Audio diartikan sebagai suara atau reproduksi suara. Gelombang suara adalah gelombang yang dihasilkan dari sebuah benda yang bergetar. Gambarnya adalah senar gitar yang dipetik, gitar akan bergetar dan getaran ini merambat di udara, atau air, atau material lainnya. Satu-satunya tempat dimana suara tak dapat

merambat adalah ruangan hampa udara. Gelombang suara ini memiliki lembah dan bukit, satu buah lembah dan bukit akan menghasilkan satu siklus atau periode. Siklus ini berlangsung berulang-ulang, yang membawa pada konsep frekuensi. Jelasnya, frekuensi adalah jumlah dari siklus yang terjadi dalam satu detik. Satuan dari frekuensi adalah Hertz atau disingkat Hz.

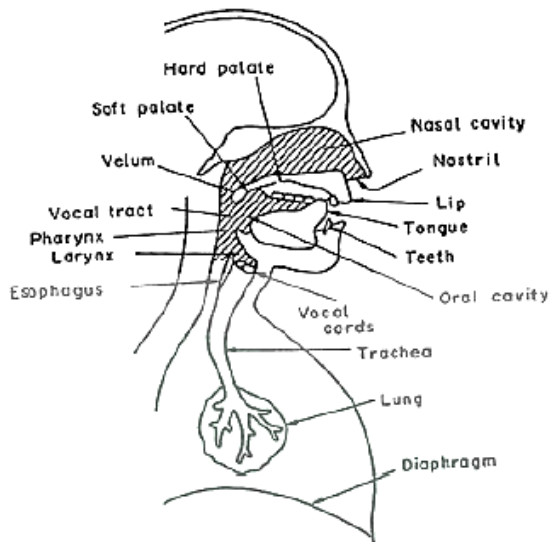
Telinga manusia dapat mendengar bunyi antara 20 Hz hingga 20 KHz (20.000 Hz) sesuai dengan batasan sinyal suara. Karena pada dasarnya sinyal suara adalah sinyal yang dapat diterima oleh telinga manusia. Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 KHz merupakan frekuensi tertinggi yang dapat didengar (Sipasulta, St and Sompie, 2014).

Merupakan bidang dalam pembelajaran mesin dan dapat diartikan sebagai "tindakan mengambil data mentah dan bertindak berdasarkan klasifikasi data". Dengan demikian, hal tersebut merupakan himpunan kaidah bagi pembelajaran terawasi (*supervised learning*).

Suatu ilmu untuk mengklasifikasikan atau menggambarkan sesuatu berdasarkan pengukuran kuantitatif fitur (ciri) atau sifat utama dari suatu obyek. Penentuan suatu obyek fisik atau kejadian ke dalam salah satu atau beberapa kategori (Putra Darma, 2010).

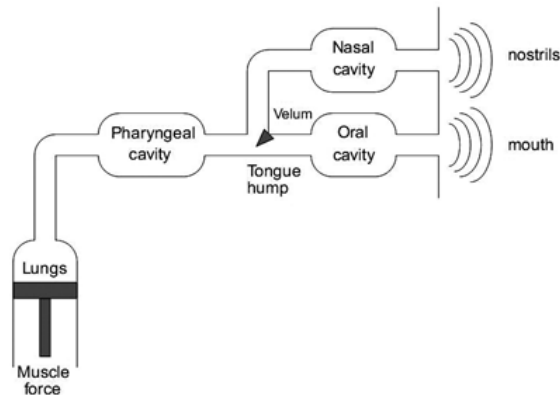
2.2.3. Proses pembentukan suara manusia

Pada manusia, proses yang mendasari terbentuknya suara terdapat tiga proses utama, antara lain proses pembangkitan sumber, proses artikulasi dan radiasi. Pada saat proses pembentukan bunyi bahas atau ucapan dimulai dengan memanfaatkan tekanan pernafasan pada rongga dada yaitu paru – paru. Pada saat manusia melakukan pernafasan tekanan udara yang ada di rongga dada yang dikeluarkan berupa arus udara. Arus udara yang dikeluarkan inilah dapat mengalami perubahan pada pita suara manusia yang ada pada pangkal tenggorokan. Pada arus udara ini dapat membuka dua pita udara yang merapat sehingga dapat menyebabkan corak bunyi tertentu. Dari gerakan membuka dan menutupnya pita suara inilah yang menjadi dasar munculnya suara yang keluar dari mulut kita dikarenakan arus udaradan udara yang berada disekitar pita suara berubah tekanan dan ikut bergetar. Berikut gambar organ – organ tubuh manusia yang berperan dalam proses pembentukan suara pada gambar 2.2.



Gambar 2.2 Organ pembentukan suara manusia (Perdana kusuma WP, 2015)

Pada proses awal paru – paru mengalami keadaan mengembang dan mengempis untuk menyedot dan mengeluarkan udara dari paru – paru. Setelah itu udara dihembuskan oleh paru – paru keluar melewati suatu daerah yang dinamakan daerah glotal. Dalam keadaan ini pita suara bervibrasi menghasilkan berbagai jenis gelombang suara. Udara kemudian melewati lorong yang dinamakan faring. Dari faring, udara melewati dua lintasan yaitu melalui hidung dan rongga mulut. Untuk organ penghasil suara manusia dapat dijelaskan pada gambar 2.3.



Gambar 2.3 Sistem produksi suara pada manusia (Perdana kusuma WP, 2015)

Manusia memiliki suara atau bunyi yang berbeda – beda yang dipengaruhi dari bentuk saluran suara yang terdiri dari rongga faring, rongga mulut dan rongga

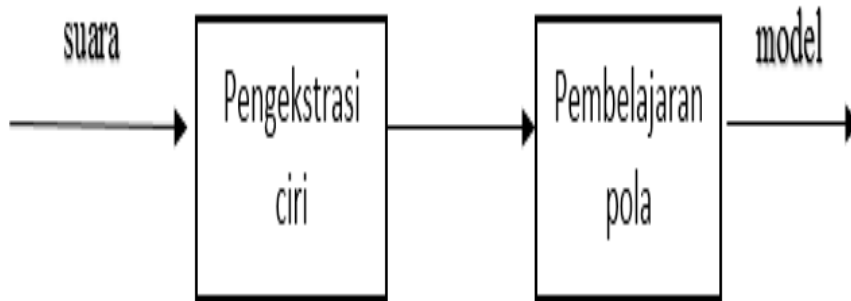
hidung. Ketika udara dari paru – paru dihembuskan, pita suara akan mengalami kondisi merapat dan merenggang. Jika kedua pita suara tersebut bergantian merapat atau merenggang dalam pembentukan suatu bunyi bahasa, maka bunyi bahasa yang dihasilkan akan terasa berat. Pada proses tersebut bunyi bahasa dinamakan bunyi bersuara (*voiced*). Akan tetapi jika bunyi bahasa merenggang dan arus udara dapat lewat dengan mudah, maka bunyi bahasa ini pada umumnya dinamakan bunyi tak bersuara (*unvoiced*).

Berdasarkan ada tidaknya rintangan terhadap arus udara, bunyi bahasa dapat dibedakan menjadi dua yaitu vokal dan konsonan. Vokal adalah bunyi udara yang arus udaranya tidak mengalami rintangan dan kualitasnya ditentukan oleh ketinggian posisi lidah, pada bagian lidah yang dinaikan dan bentuk bibir. Sedangkan konsonan adalah bunyi bahasa yang arus udaranya mengalami rintangan dan kualitasnya ditentukan oleh keadaan pita suara, penyentuhan atau pendekatan berbagai alat ucap dan cara alat ucap itu bersentuhan.

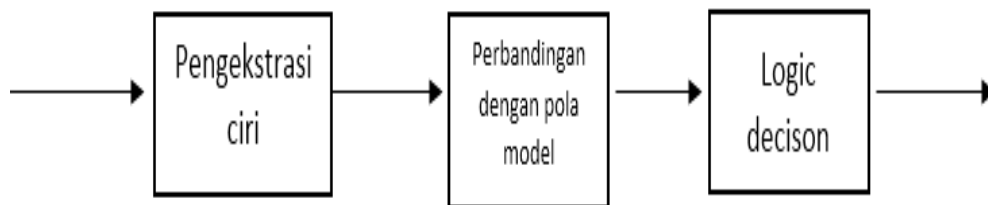
2.2.4. Pengenalan Suara

Pengenalan suara merupakan salah satu upaya agar suara dapat dikenali atau diidentifikasi sehingga dapat dimanfaatkan. Pengenalan suara dapat dibedakan kedalam tiga bentuk pendekatan, yaitu pendekatan akustik - fonetik (*The acoustic – phonetic approach*), pendekatan kecerdasan buatan (*The Artificial Intellegence approach*), dan pendekatan pengenalan pola (*The Pattern Recognition Approach*) (Perdana Kusma WP,2015).

Pendekatan pengenalan pola terdiri dari dua langkah yaitu pembelajaran pola dan pengenalan pola melalui perbandingan pola. Tahap perbandingan pola adalah tahap bagi ucapan yang akan dikenali, dibandingkan polanya dengan setiap kemungkinan pola yang telah dipelajari dalam fase pembelajaran, untuk kemudian diklasifikasi dengan pola terbaik yang cocok. Blok diagram pembelajaran pola pada pengenalan suara ditunjukkan pada gambar 2.4 dan 2.5.



Gambar 2.4 Blok diagram pembelajaran pola (Perdana kusuma WP, 2015)



Gambar 2.5 Blok diagram pengenalan suara (Perdana kusuma WP, 2015)

Berikut ini merupakan penjelasan dari masing – masing blok :

1. Pengekstrasi ciri
Merupakan proses mendapatkan sederetan besaran pada bagian sinyal masukan untuk menetapkan pola pembelajaran atau pola uji. Pada sinyal suara, ciri – ciri besaran biasanya merupakan keluaran dari beberapa bentuk teknik analisis spektrum seperti *Fast Fourier Transform* (FFT).
2. Pembelajaran pola
Satu atau lebih pola pembelajaran yang berhubungan dengan bunyi suara dari kelas yang sama, digunakan untuk membuat pola representatif dari ciri – ciri kelas tersebut. Hasilnya yang disebut dengan pola referensi, dapat menjadi sebuah model yang mempunyai karakteristik bentuk statistik dari ciri – ciri pola referensi.
3. Perbandingan dengan pola model

Pola uji yang akan dikenali, dibandingkan dengan setiap kelas pola referensi. Kesamaan besaran antara pola uji dengan setiap pola referensi akan dihitung.

4. *Logic decision*

Menentukan kelas pola referensi mana yang paling cocok untuk pola uji berdasarkan klasifikasi pola.

Pengenalan suara atau ucapan (*speech recognition*) merupakan suatu teknik yang memungkinkan sistem komputer untuk menerima input berupa kata yang diucapkan. Kata –kata tersebut diubah bentuknya menjadi sinyal digital dengan cara merubah gelombang suara menjadi sekumpulan angka lalu disesuaikan dengan kode – kode tertentu dan dicocokkan dengan suatu pola yang tersimpan dalam suatu perangkat. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi.

Teknologi pengenalan suara berkembang dengan begitu pesat, beberapa jenis diantaranya yaitu :

1. Sintesis suara (*speech synthesis*) : dari teks ke suara.
2. Pengenalan pembicara (*speaker recognition*) : dari suara ke identitas pembicara.
3. Pendiktean (*dictation*) : dari suara ke teks.
4. Rangkuman suara (*speech summarization*) : dari suara ke teks sederhana.
5. Pengkategorian suara (*speech categorization*) : dari suara ke label kelas .
6. Pengertian suara (*speech understanding*) : dari suara ke representasi makna suara.
7. Pemrosesan dialog (*dialog processing*) : dari suara ke makna suara yang interaktif.
8. Penerjemah suara (*speech translation*) : dari suara ke suara dalam bahasa lain.

Kegunaan aplikasi pengenalan suara yaitu :

1. Pada bidang komunikasi yaitu sebagai komando suara pada komputer untuk melakukan perintah seperti membuka, menyimpan dan menutup file.
2. Pada bidang kesehatan alat pengenal ucapan digunakan untuk membantu para penyandang cacat agar dapat beraktifitas dengan memerintahkan alat – alat bantu melalui suaranya.

3. Pada bidang militer yaitu digunakan pada pengatur lalu lintas udara yang dikenal dengan *Air Traffic Controller* (ATC). Alat ini dipakai untuk memberi keterangan mengenai keadaan lalu lintas udara seperti radar, cuaca, dan navigasi dengan kata lain sebagai pengganti operator yang memberikan informasi kepada pilot dengan cara berdialog.
4. Secara umum pengenalan ucapan banyak digunakan karena mudah yakni hanya menggunakan suara dan cepat prosesnya.

Kekurangan pengenalan ucapan yaitu :

- a. Rawan terhadap gangguan sinyal suara lain terutama di tempat yang ramai.
- b. Kata – kata yang diucapkan dapat sulit dikenali karena cara pengucapan yang berbeda walaupun oleh pembicara yang sama.
- c. Bahasa lisan seringkali diucapkan tidak sesuai dengan kaidah tata bahasa yang baku.
- d. Jumlah kata yang dapat dikenali masih terbatas.

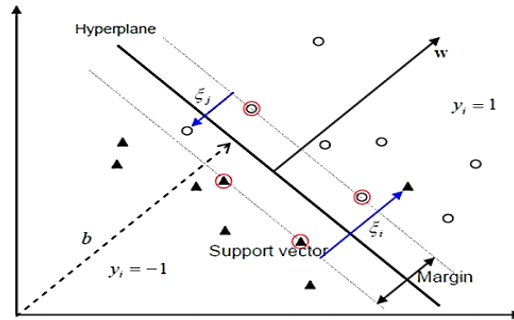
2.2.5. Metode *Fast Fourier Transform* (FFT)

Dalam pertengahan tahun 1960, J. W. Cooley dan J. W. Tukey, berhasil merumuskan suatu teknik penghitungan *Fourier Transform* yang efisien. *Fast Fourier Transform* (FFT) menjadi penting untuk bermacam – macam aplikasi, dari pengolahan sinyal digital dan memecahkan persamaan *diferensial parsial* menjadi algoritma – algoritma untuk penggandaan bilangan *integer* dalam jumlah yang banyak. Algoritma FFT merupakan penyederhanaan dari *Diskrit Fourier Transform* (DFT). Dengan metode FFT, laju dari komputasi perhitungan *Fourier Transform* dapat ditingkatkan, komputasi DFT adalah komputasi yang memerlukan waktu untuk proses looping dan memerlukan banyak memori. Dengan menerapkan FFT, perhitungan DFT dapat dipersingkat, dalam hal ini proses looping dapat direduksi. DFT memiliki persyaratan jumlah data harus merupakan 2^n untuk $n=0,1,2,3,\dots$ waktu DFT memiliki kompleksitas N^2 sedangkan FFT memiliki kompleksitas $Np/2$ dengan $p=2 \log N$, sehingga FFT lebih cepat daripada DFT dengan rasio kecepatan FFT terhadap DFT (Anam, khoirul 2013).

2.2.6. Metode *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah sistem pembelajaran untuk mengklasifikasikan data menjadi 2 kelompok yaitu menggunakan fungsi-fungsi

linear dan non-linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi (Ikra Dewantara, 2013). Klasifikasi SVM dilakukan dengan cara menemukan hyperplane terbaik sehingga diperoleh ukuran margin yang maksimal. Margin adalah jarak antara hyperplane tersebut dengan titik terdekat dari masing-masing kelas. Titik yang paling dekat ini disebut dengan support vector pada gambar 2.6.



Gambar 2.6 Ilustrasi *Support Vector Machine* (SVM) linier (Ikra Dewantara, 2013)

Gambar 2. mengilustrasikan dua kelas yang dapat dipisahkan secara linier menggunakan sepasang bidang batas yang sejajar. Bidang batas pertama membatasi kelas pertama ($y_i = 1$), dan bidang batas kedua membatasi kelas kedua ($y_i = -1$). Persamaan bidang pembatas jika data terpisah secara linier dapat dihitung menggunakan perkalian vektor antara vektor bobot dengan data set seperti disajikan pada Persamaan (1) dan (2) :

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

$$w \cdot x_i + \leq b - \text{ untuk kelas 2 } y_i = -1 \quad (2.2)$$

dimana :

x_i = data set

y_i = kelas dari data x_i

w = vektor bobot yang tegak lurus terhadap hyperplane .

b = menentukan fungsi pemisah relatif terhadap titik asal.

Bidang pemisah terbaik adalah bidang pemisah yang memiliki margin (jarak antara dua bidang pembatas) maksimal atau berada di tengah-tengah kedua kelas yang berbeda. Nilai margin antara dua bidang pembatas adalah $m = \frac{2}{\|w\|}$. Margin

maksimal diperoleh dengan menggunakan fungsi Lagrangian pada persamaan (3):

$$\min_{w,b,L_p} L_p(w,b,a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^1 (a_i y_i ((w \cdot x + b) - 1)) + \sum_{i=1}^1 a_i \quad (2.3)$$

Vektor w sering kali bernilai sangat besar bahkan tidak terhingga, tetapi untuk nilai terhingga. Persamaan *Lagrange primal problem* perlu diubah ke dalam *Lagrange dual problem* seperti Persamaan (4) :

$$L_D(a) = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{j=1}^1 a_i a_j y_i y_j x_i x_j \quad (2.4)$$

Karena $\min_{w,b} L_p = \max_a L_d$. Sehingga solusi pencarian bidang pemisah terbaik *disajikan dalam Persamaan* (5) dan (6) :

$$\text{Max}_a L = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{i=1}^1 \sum_{j=1}^1 a_i a_j y_i y_j x_i x_j \quad (2.5)$$

$$\text{s. t } \sum_{i=1}^1 a_i y_i = 0 \geq 0 \quad (2.6)$$

Persamaan (6) akan menghasilkan nilai a_i untuk setiap data pemodelan. Nilai tersebut digunakan untuk menentukan bobot (w). Data yang memiliki nilai $a_i \geq 0$ adalah *support vector*, sedangkan sisanya dianggap bukan *support vector*. Setelah a_i ditemukan, maka kelas dari data pengujian x_i dapat ditentukan berdasarkan Persamaan (7) :

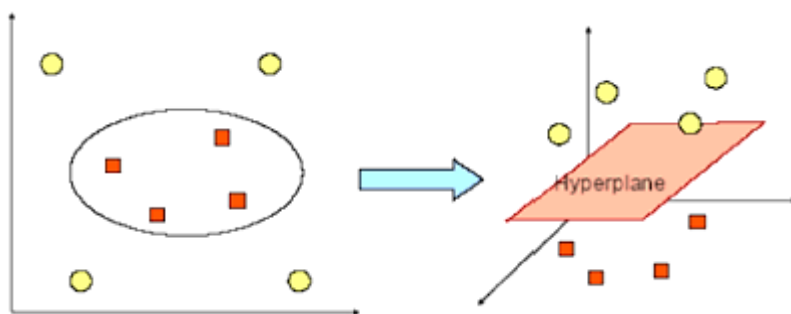
$$f(x_d) = \sum_{i=1}^{NS} a_i y_i x_i x_d + b \quad (2.7)$$

dimana:

x_i = *support vector*

NS = jumlah *support vector*

x_d = data yang akan diklasifikasikan



Gambar 2.7 Ilustrasi *Support Vector Machine* (SVM) non- linear
(Ikra Dewantara, 2013)

Gambar 2.7 mengilustrasikan di ruang vektor baru tersebut SVM mencari *hyperplane* terbaik untuk memisahkan dua kelas secara *linear*. Pencarian *hyperplane* bergantung kepada *dot product* dari data yang dipetakan pada ruang berdimensi tinggi, yaitu $\phi(u) \phi(v)$. Perhitungan transformasi $\phi(u)$ ini sangat sulit, namun dapat digantikan dengan fungsi yang disebut dengan *kernel* sesuai pada Persamaan (8) :

$$K(u,v) = \phi(u) \cdot \phi(v) \quad (2.8)$$

Sehingga jika disubstitusi ke dalam persamaan *Lagrangian* menjadi seperti Persamaan 8 dengan fungsi yang dihasilkan pada Persamaan (9) :

$$L = \sum_{i=1}^1 a_i - \frac{1}{2} \sum_{j=1}^1 a_j y_j K(u,v) \quad (2.9)$$

$$F(x_d) = \sum_{i=1}^{NS} a_i y_i K(u,v) + b$$

dimana:

u = data latih

v = kelas pada data latih

d dan γ adalah parameter *kernel*.

Pemodelan pada *kernel linear* dan *polynomial* merupakan perkalian vektor antara data latih dan kelas yang bersesuaian. Adapun data latih (u) harus ditranspose terlebih dahulu menjadi (u^T) agar dimensi datanya sesuai.

Halaman Ini Sengaja Dikosongkan