

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 Proposal Skripsi**

Skripsi adalah istilah yang digunakan di Indonesia untuk mengilustrasikan suatu karya tulis ilmiah berupa paparan tulisan hasil penelitian sarjana S1 yang membahas suatu permasalahan/fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah-kaidah yang berlaku [1]. Proposal skripsi sendiri merupakan Suatu permintaan, pengajuan ide kepada pihak lain untuk mendapatkan perizinan topik, tempat penelitian, dan lain sebagainya yang diajukan oleh seorang mahasiswa dalam usahanya untuk melakukan skripsi sebagai syarat kelulusannya sebagai sarjana [4].

Fungsi porposal skripsi penting untuk seorang mahasiswa tingkat akhir yang akan melakukan penelitian sebagai syarat kelulusannya, antara lain [4]:

1. Untuk mengajukan persetujuan topik tugas akhir.
2. Untuk mengajukan tempat penelitian.
3. Untuk mendapatkan perizinan yang menunjang penelitian seperti akses data, melakukan observasi, melakukan wawancara, dan lain sebagainya di tempat penelitian.

Dalam menyusun proposal skripsi harus memperhatikan sistematika penulisannya, sistematika penulisan sendiri disusun untuk memberikan gambaran umum tentang penelitian yang akan dilakukan [4]. Sistematika penulisan proposal tergantung pada jenis penelitiannya, tapi pada umumnya isi dari sistematika penulisan seperti dibawah ini:

1. Halaman Judul

Berisi judul dari proposal skripsi, logo universitas, dan keterangan fakultas, program studi dan nama universitas beserta tahun penyusunan. Halaman judul menggambarkan bagaimana isi dari sebuah skripsi yang dikerjakan terutama terlihat dari judul yang tertera.

## 2. Daftar Isi

Berisi daftar menu dan sub menu yang ada di proposal skripsi. Daftar isi dapat mempermudah untuk menuju ke poin tertentu yang sudah dialamatkan dengan halaman tertentu.

## 3. Daftar Tabel

Berisi daftar tabel yang digunakan pada proposal skripsi. Mempermudah untuk melihat semua tabel yang tersedia serta mempermudah mengakses tabel yang tersedia dengan keterangan alamat halaman yang ada.

## 4. Daftar Gambar

Berisi daftar gambar yang terdapat pada proposal skripsi. Mempermudah untuk melihat semua gambar yang tersedia serta mempermudah mengakses gambar yang tersedia dengan keterangan alamat halaman yang ada.

## 5. Latar Belakang masalah

Berisi latar belakang dalam pembuatan penelitian. Serta memberikan gambaran tentang skripsi yang dikerjakan dari peneliti skripsi.

## 6. Identifikasi Masalah

Berisi identifikasi masalah dalam penelitian. Serta menjabarkan masalah yang dihadapi untuk diselesaikan oleh peneliti skripsi.

## 7. Maksud dan Tujuan

Berisi maksud dan dan tujuan diadakannya penelitian tersebut. Sehingga memperjelas alasan peneliti untuk melakukan penelitian skripsi tersebut.

## 8. Batasan Masalah

Berisi batasan dalam melakukan penelitian tersebut. Dengan batasan masalah peneliti menjabarkan batasannya melakukan penelitian sehingga menjelaskan sejauh mana penelitian dikaji.

## 9. Metodologi Penelitian

Berisi Metodologi penelitian yang akan dipakai, serta menjelaskan bagaimana alur penelitian dilakukan.

## 10. Deskripsi Umum Sistem

Berisi gambaran sistem yang akan dibangun. sehingga memperjelas sistem yang akan dibangun sehingga pembaca dapat mengerti alur sistem yang akan dibangun.

#### 11. Review Literatur

Berisi review jurnal yang dipakai sebagai pembanding penelitian. Serta memperlihatkan apa peningkatan yang peneliti ambil dari skripsi sebelumnya.

#### 12. Jadwal dan Tempat Penelitian

Berisi waktu perkiraan pengerjaan penelitian dan tempat penelitian berlangsung. Sehingga dengan adanya jadwal, sebagai peneliti bisa memperkirakan waktu kerja serta batas waktu untuk menyelesaikan penelitian skripsi tersebut.

#### 13. Sistematika Penulisan

Berisi sistematika dalam penulisan skripsi, dan dalam penulisannya sudah diatur sebagai gambaran umum penulisan skripsi di Universitas Komputer Indonesia.

#### 14. Daftar Pustaka

Berisi seluruh kepustakaan yang digunakan dalam penyusunan proposal. Dapat membantu para peneliti lain yang hendak mengambil tema skripsi yang menyerupai untuk mencari jurnal atau buku yang bersangkutan dengan penelitian yang diambil.

Kelompok keilmuan Teknik Informatika UNIKOM mengadopsi Computing Curricula 2005[26], yaitu :

##### 1. Kelompok A (Sistem Informasi), yang mencakup di antaranya:

###### a) Sistem Informasi Geografis (SIG)

SIG menyajikan informasi dalam bentuk grafis dengan menggunakan peta sebagai antarmuka, saat ini banyak digunakan untuk perencanaan, pelaksanaan, dan pengendalian yang berkaitan dengan wilayah geografis.

Contoh yang termasuk dalam topik ini diantaranya :

- a. Sistem Informasi Geografis Untuk Menunjang Rencana Strategis Penanggulangan Tumpahan Minyak Di Selat Malaka
- b. Sistem Penentuan Lokasi untuk Membuka Cabang Baru Toko Kue X Menggunakan Sistem Informasi Geografis
- c. Sistem Informasi Geografis Asrama Universitas ABC
- d. Sistem Informasi Geografis Pemantauan Penyebaran Penyakit Demam Berdarah Di Kota XYZ

- e. Sistem Informasi Geografis Penempatan BTS Provider XYZ Di Kabupaten X
  - f. Dan lain-lain.
- b) Supply Chain Management (SCM)
- SCM adalah koordinasi dari bahan, informasi dan arus keuangan antara perusahaan yang berpartisipasi. Contoh yang termasuk dalam topik ini diantaranya :
- a. Sistem Pengendalian Distribusi Sepatu Menggunakan Pendekatan Supply chain Management di PT. ABC
  - b. Supply Chain Management Produksi Obat ABC di PT. XYZ
  - c. Dan lain-lain
- c) Sistem Informasi Manajemen (SIM)
- SIM adalah sistem perencanaan bagian dari pengendalian internal suatu bisnis yang meliputi pemanfaatan manusia, dokumen, teknologi, dan prosedur oleh pihak manajerial untuk memecahkan masalah bisnis seperti biaya produk, layanan, atau suatu strategi bisnis. Sistem informasi manajemen dibedakan dengan sistem informasi biasa karena SIM digunakan untuk menganalisis sistem informasi lain yang diterapkan pada aktivitas operasional organisasi. SIM dibangun agar organisasi memiliki informasi yang bermanfaat dalam pembuatan keputusan manajemen, baik yang menyangkut keputusan-keputusan rutin maupun keputusan-keputusan yang strategis. Contoh yang termasuk dalam topik ini diantaranya :
- a. Sistem Informasi Manajemen Bagian Produksi PT. XYZ
  - b. Sistem Informasi Manajemen Simpan Pinjam di Bank Y
  - b. Dan lain-lain
- d) Sistem Informasi Eksekutif (EIS)
- EIS merupakan sebuah sistem berbasis komputer yang menyediakan data & informasi (baik internal maupun eksternal) yang dibutuhkan oleh top eksekutif, dimana informasi disediakan dengan cepat & tepat serta disajikan dalam bentuk laporan-laporan yang didukung grafik ataupun kemampuan drill-down. Pada umumnya, informasi tersebut akan digunakan oleh top eksekutif untuk memantau kondisi bisnis perusahaan secara umum atau hasil

sebuah proses bisnis secara khusus. Oleh karena itu, penelitian skripsi dengan tema EIS merupakan penelitian yang berada dalam lingkup pembangunan EIS ataupun pengembangan EIS. Contoh yang termasuk dalam topik ini diantaranya:

- a. Sistem Informasi Eksekutif PT. ABC
  - b. Sistem Informasi Eksekutif Departemen Logistik PT. ABC
  - c. Sistem Informasi Eksekutif Departemen Logistik PT. ABC untuk Menganalisis Total Biaya Penyimpanan Optimum
  - c. Sistem Informasi Eksekutif PT. ABC untuk Memonitoring Kinerja Departemen Logistik.
  - d. Sistem Informasi Eksekutif PT. ABC untuk Memonitoring Kinerja Perusahaan
  - e. Dan lain lain
- e) Enterprise Architecture (EA)

EA merupakan sebuah gambaran As-Is & To-Be suatu enterprise yang berkaitan dengan mengorganisasikan seluruh proses bisnis enterprise, informasi yang dibutuhkan serta teknologi pendukungnya. Oleh karena itu, penelitian skripsi dengan tema ini merupakan penelitian yang menganalisis serta menyelaraskan antara penerapan sistem informasi dengan kebutuhan enterprise (organisasi), sehingga terbentuk suatu rancangan strategis dan komponen arsitektural (seperti arsitektur bisnis, arsitektur data, arsitektur informasi & arsitektur teknologi). Contoh yang termasuk dalam topik ini diantaranya:

- a. Perancangan Enterprise Arsitektur Planning di PT.ABC
- b. Perancangan Enterprise Arsitektur Planning (EAP) pada Proses Manajemen Aset dengan Zachman Framework (Studi Kasus Divisi Manajemen Fasilitas PT. ABC)
- c. Perancangan Enterprise Arsitektur Menggunakan Togaf Versi 9 (Studi Kasus : Dewan Kehormatan Penyelenggaraan Pemilu (DKPP))
- d. Pemodelan Enterprise Arsitektur Zachman Framework pada Sistem Informasi Fakultas Teknologi Informasi Universitas ABC
- e. Dan lain lain

f) Knowledge Management System (KMS)

Knowledge Management (KM) berisi serangkaian strategi, sistem dan teknik yang digunakan oleh individu, team dan organisasi untuk mengelola 'knowledge'. Sesuai dengan sifat 'knowledge' yang demikian subjective, validasi sebuah 'individual knowledge' menjadi 'collective knowledge' hanya bisa tercapai jika terbangun proses knowledge sharing tukar menukar individual knowledge. Collective knowledge yang tervalidasi dan terdokumentasi adalah sumber kekuatan organisasi. Contoh yang termasuk dalam topik ini diantaranya:

- a. Sistem Pengelolaan Pengetahuan Untuk Meningkatkan Kreasi Dan Inovasi Perusahaan
- b. Pengelolaan Materi Ajar dengan Pendekatan Knowledge Management System Di Instansi ABC
- c. Dan lain-lain

g) Customer Relationship Management (CRM)

CRM adalah sebuah sistem informasi yang terintegrasi yang digunakan untuk merencanakan, menjadwalkan, dan mengendalikan aktivitas-aktivitas pra penjualan dan pasca penjualan dalam sebuah organisasi. CRM melingkupi semua aspek yang berhubungan dengan calon pelanggan dan pelanggan saat ini, termasuk di dalamnya adalah pusat panggilan (call center), tenaga penjualan (sales force), pemasaran, dukungan teknis (technical support), layanan lapangan dan strategi organisasi atau perusahaan yang berfokus pada pembangunan dan pemeliharaan hubungan dengan pelanggan. Contoh yang termasuk dalam topik ini diantaranya:

- a. Pengaruh Penerapan Customer Relationship Management (CRM) Terhadap Loyalitas Pelanggan PT. Bank Mandiri Tbk, Cabang Braga Bandung.
- b. Dan lain-lain

h) Sistem Informasi Manajemen Proyek

Manajemen proyek merupakan aplikasi atau implementasi dari pengetahuan, keterampilan, perangkat dan teknik pada suatu aktivitas proyek untuk memenuhi kebutuhan atau tujuan suatu proyek. Perencanaan,

pemantauan dan pengontrolan terhadap aspek yang terdapat dalam sebuah proyek, serta motivasi yang ada didalamnya untuk mencapai tujuan proyek dengan waktu, biaya, kualitas dan performansi yang telah ditentukan. Sekumpulan penugasan/pekerjaan, teknik, serta perangkat yang diaplikasikan selama eksekusi atau pelaksanaan proyek. Contoh yang termasuk dalam topik ini diantaranya:

- a. Sistem Informasi Manajemen Proyek PT. ABC
- b. Manajemen Resiko Proyek PT.XYZ
- c. Sistem Informasi Monitoring Proyek PT.ABC

2. Kelompok B (RPL, Data dan BI), yang mencakup di antaranya:

a) Rekayasa Perangkat Lunak

Bidang keilmuan Teknik Informatika yang membahas tentang konsep keilmuan yang digunakan dalam membangun/menilai sebuah perangkat lunak. Kajian untuk bidang ini diantaranya :

- a. Coding Assesment
- b. Pembangunan Framework
- c. Design Pattern Finding
- d. Code Refinement
- e. Pembangunan Class Library
- f. Format Method
- g. Human Computer Interaction
- h. Software Quality
- i. Software Matrics
- j. Software Testing
- k. Software Migration
- l. Software Refinement

b) Rekayasa Data

Bidang keilmuan Teknik Informatika yang membahas model data, implementasi teknologi basis data dan terapannya dalam kehidupan sehari-hari. Kajian untuk bidang ini diantaranya :

- a. Data Mining
- b. Data warehouse

- c. Clustering Database
  - d. Big Data
  - e. Web Mining/Text Mining
  - f. Basis Data Terdistribusi
  - g. NoSQL Performance
  - h. Data SpatioTemporal
  - i. Data Spatial
  - j. Query Processing and Optimization
  - k. Data Master Management
- c) Business Intelligence
- Aplikasi, teknologi, dan metodologi untuk mengumpulkan, menyimpan dan menyediakan akses terhadap data untuk menolong enterprise mengembangkan bisnisnya.
3. Kelompok C (Jaringan Komputer dan Multimedia), yang mencakup di antaranya:
- a) Multimedia
 

Terkait dengan pengolahan data untuk menghasilkan informasi dengan media-media tertentu – atau gabungan media – menghasilkan nilai informasi yang lebih tinggi. Kajian untuk bidang ini diantaranya :

    - a. Voice Technology. Speech Recognition, Speaker Recognition
    - b. Video Recognition. Bisa berupa pengenalan wajah/objek pada video
    - c. Multimedia sebagai sarana pemberi informasi
    - d. Virtual Museum, Virtual Reality dan Augmented Reality
    - e. Chatbot Multimedia
    - f. Game (Algoritma/Metode)
    - g. Kajian. Analisis Metode, Prototipe, Implementasi
    - h. Platform. Desktop, WEB, Mobile
  - b) Jaringan Komputer dan Internet
 

Membangun atau mengembangkan sistem dan konten yang terdapat dilingkungan jaringan komputer (LAN/WAN). Bukan untuk membangun jaringan (LAN/WAN). Bentuk pembahasan dalam skripsi untuk bidang ini dapat berupa :

- a. Pembangunan Sistem / Aplikasi terkait dengan Sistem Jaringan
- b. Komputer (Desktop, Web atau Mobile)
- c. Pembangunan Sistem dengan Konfigurasi Tools – Untuk Menghasilkan Sistem Baru.
- d. Menganalisis Performansi – Jaringan Komputer
- e. Menganalisis Quality Of Services (QoS) Pada Sistem Jaringan
- f. Komputer

Termasuk didalamnya Teknologi Distributed Computing (Grid, Paralel Computing, Pervasive Computing) hingga teknologi Cloud Computing.

Kajian untuk bidang ini diantaranya :

- a. Aplikasi dengan protokol multimedia jaringan untuk menyediakan layanan multimedia di dalam jaringan (Video on Demand (VOD), Live Streaming (Audio, Video). Fungsi komunikasi, entertainment, edukasi.
  - b. Pembangunan VOIP (Voice Over Internet protocol), VPN (Virtual Private Network). Metode, dan sistem.
  - c. Kajian-Kajian Keamanan Pada jaringan Komputer dan Internet
  - d. Metode/Algoritma pada Sistem Komputer Terdistribusi, Grid/Paralel Sistem
  - e. Cloud Computing – Aplikasi, Sekuriti dan Arsitektur.
  - f. Kajian IPv6, Metode, Migrasi Plan dsb
- c) Keamanan Sistem

Membahas kajian keamanan dari sistem disertai solusinya. Kajian untuk bidang ini diantaranya :

- a. Kajian Keamanan Sistem
    - Dilingkungan Standalone, Dilingkungan Jaringan + Internet
  - b. Platform
    - Desktop, Web, Mobile
  - c. Tema Keamanan Sistem
    - VPN, Firewall, Proxy, Kriptografi (Metode/Algoritma), Steganografi (Metode/Algoritma), Digital Signature, Secure Coding - Kajian Ancaman-ancaman Keamanan disertai Solusinya.
- d) Teknologi IT

Pemanfaatan teknologi-teknologi IT pada berbagai bidang/aplikasi. Kajian untuk bidang ini diantaranya :

a. QRCode

Untuk berbagai keperluan basis WEB dan mobile(HTML 5) dengan akses data via mobile menggunakan data asli di Web

b. RFID

Untuk berbagai keperluan, Contoh Dapat RFID sistem Untuk Unikom atau RFID untuk Stadion, Pembayaran Bus Transjakarta/Trans bandung dsb)

c. NFC

Transfer pulsa, data dsb (teknologi baru, terbatas pada HP yang mendukung NFC)

d. Internet of Things

Smart System (Smart City, Smart Transportation, Smart Traffic, Smart Farming, Smart Fishing dsb) Dapat diintegrasikan dengan Microcontroller, Micro PC (Arduino), sensor dsb. Namun tetap inti tetap aplikasi software sistemnya Cakupan luas, Desa, rumah sakit, Kantor, UKM, Puskesmas, pendidikan, bengkel, Pesantren, Masjid dsb.

4. Kelompok D (Teknologi Internet dan Mobile), yang mencakup di antaranya:

a) E-Learning

Pembelajaran yang diselenggarakan melalui media elektronik. Audio, Video, Komputer, Tablet, Perangkat Mobile, Blog, Webcam, Smart Whiteboards, Virtual Classroom, atau Penggabungan Teknologi. Contoh Kajian untuk bidang ini diantaranya :

a. Pembangunan Perangkat Lunak Pembelajaran (Fisika, Matematika, Membaca Al-Quran, Tulisan Sunda, dll) (CDInteraktif, Situs)

b. Pembangunan Perangkat Lunak Untuk Pembuatan Materi Pembelajaran Online.

c. Situs Agregator Konten Pembelajaran di Departemen Pendidikan

b) Learning Management System (LMS)

Perangkat lunak yang digunakan untuk mengirimkan/mendistribusikan, menelusuri dan mengatur pembelajaran. Dengan sub sistem Pendataan peserta, materi, instruktur, pemberian tugas, kehadiran, pengumuman, penilaian, latihan, diskusi dan pelacakan progress pembelajaran. Contoh Kajian untuk bidang ini diantaranya :

- a. Pengembangan Learning Management System untuk SMA XYZ untuk pemeriksaan ujian essay secara otomatis.
  - b. Pengembangan Learning Management System untuk Lembaga Kursus XYZ untuk Penjadwalan Kursus yang optimal dengan algoritma ABC.
  - c. Pengembangan Learning Management Sistem berbasis Cloud
- c) WEB/Internet Technology

Pemanfaatan teknologi web/internet sesuai kebutuhan. Teknologi yang bisa digunakan diantaranya Cloud Computing, Web 2.0, Web Service, Networking, HTML5, XML, CSS, JSON. Contoh Kajian untuk bidang ini diantaranya :

- a. Kolaborasi Sistem Informasi Dampak Bencana dan Penanggulangannya (Kombinasi dari Online Map, GPS, Mobile)
  - b. Pemanfaatan Google Maps dalam Sistem Monitoring Persediaan Pangan di Badan Logistik Jawa Barat.
  - c. Monitoring kondisi lalu lintas di kota Bandung berdasarkan informasi dari media sosial. (Web Analytic, Media Sosial, Online Map).
  - d. Pembangunan Infrastruktur sistem E-Desa berbasis cloud computing.
  - e. Pembangunan Sistem Layanan File Sharing Khusus Untuk Materi Pembelajaran
- d) Pengembangan E-Commerce

Tidak membuat situs e-commerce dari nol, tetapi pengembangan system e-Commerce yang sudah ada. Sistem yang sudah ada boleh dari system apa pun. Teknologi yang bisa ditambahkan adalah Data Mining (Association Rule, Classification, dll), Web Analytic, Supply Chain Management (SCM), Customer Relationship Management (CRM), Online Map, Online Payment Gateway, Tracking, Social Media. Contoh Kajian untuk bidang ini diantaranya :

- a. Pemanfaatan Association Rule Untuk Promosi Produk di Toko Online XYZ (Data Mining di data toko online)
  - b. Integrasi Customer Relationship Management (CRM) di Toko Online XYZ
  - c. Optimalisasi Penjadwalan Produksi Berdasarkan Pesanan Online di PT. XYZ (termasuk Supply Chain Management, Resource Management)
  - d. Pembangunan Infrastruktur Sistem Integrator Data Produk di Factory Outlet Bandung. (Data mining, Online Maps, Web Service)
  - e. Sistem Monitoring Trend di Social Media untuk Memprediksi Produk yang Dibutuhkan Masyarakat (Targeted Marketing, Web Analysis, Data Mining)
- e) Mobile Technology
- Pemanfaatan teknologi mobile di berbagai bidang. Teknologi yang bias anda manfaatkan diantaranya Location Based Service (LBS) / GPS, Web Service, Image Processing / Augmented Reality, Data Mining, Sensor. Contoh Kajian untuk bidang ini diantaranya :
- a. Pemanfaatan GPS di perusahaan Taxi XYZ (GPS, Web, Database, Mobile Application, Web Service)
  - b. Pemanfaatan Teknologi Augmented Reality sebagai Virtual Guide Lokasi Bersejarah Kota Bandung.
  - c. Aplikasi Wisata Kuliner di Kota Bandung (Data Aggregator, Data Mining, Augmented Reality)
  - d. Pembangunan Aplikasi Pelaporan Distribusi Bantuan untuk Daerah Bencana Alam Berbasis Mobile
  - e. Sistem Peringatan Dini Banjir Berbasis Mobile (Sistem Notifikasi, Web Service, GPS)
5. Kelompok E (Computer Science / CS) kajian teoritis dan algoritmik disertai aplikasinya, yang mencakup di antaranya:
- a) Artificial Intelligence
- Kajian yang terkait adaptasi kemampuan manusia atau hewan ke dalam sistem komputer. Contoh Kajian untuk bidang ini diantaranya :
- a. Perceptual Computing

- b. Machine Learning (Statistical Machine Learning, Machine Learning in Robot)
  - c. Optimization case
- b) Intelligent Tutoring
 

Kajian tentang sistem yang membantu proses pembelajaran dan evaluasinya. Contoh Kajian untuk bidang ini diantaranya :

  - a. Essay Grading
  - b. Question Generation
  - c. Programming learning (error-checking, suggestions, peer grading, style analysis)
- c) Information Retrieval
 

Kajian yang mempelajari pencarian informasi tertentu yang relevan dengan kebutuhan (query) berdasarkan sekumpulan sumber informasi. Contoh Kajian untuk bidang ini diantaranya :

  - a. Indexing
  - b. Search Engine
- d) Natural Language Processing
 

Kajian yang terkait pengenalan, pembelajaran suatu bahasa manusia supaya dapat dipahami komputer. Contoh Kajian untuk bidang ini diantaranya :

  - a. Information Extraction,
  - b. Question Answering,
  - c. Sentiment & Social Meaning,
  - d. Parsing & Tagging,
  - e. Machine Translation,
  - f. Dialog & Speech Processing,
  - g. Multilingual NLP,
  - h. Linguistic Structure,
  - i. Summarization
- e) Programming Language & Environments
 

Kajian yang terkait pembangunan sistem kompilasi Bahasa pemrograman dan elemen-elemen di dalamnya. Contoh Kajian untuk bidang ini diantaranya :

- a. Automatic Analyzer of Correctness Program,
  - b. Languages Comparison,
  - c. Software Support for Languages (Compiler),
  - d. Program Optimization,
  - e. Formal Reasoning,
  - f. Automated Debugging (Finding Bug Fixes, Bugs Detection)
- f) Image Recognition
- Kajian yang terkait pengenalan pola citra (gambar 2 dimensi). Contoh Kajian untuk bidang ini diantaranya :
- a. Pengenalan Pola Citra,
  - b. Pre-processing (Thinning, Edge Detection, Ekstraksi Ciri)
- g) Voice Recognition
- Kajian yang terkait pengenalan pola suara atau bunyi. Contoh Kajian untuk bidang ini diantaranya :
- a. Speaker Recognition,
  - b. Speech Recognition
- h) Teori Komputasi
- Kajian yang terkait pencarian solusi atau pembuatan model tertentu untuk menyelesaikan masalah-masalah komputasi teoritik. Contoh Kajian untuk bidang ini diantaranya :
- a. Teori Kompleksitas Komputasi (Masalah  $P = NP$ ),
  - b. Kriptografi,
  - c. Teori Quantum Computing,
  - d. Teori Automata

## 2.2 Pengertian Implementasi

Menurut Rimaru (dalam Rita Prima Bendriyanti dan Leni Natalia Zulita, 2012), implementasi merupakan suatu proses mendapatkan suatu hasil yang sesuai dengan tujuan atau sasaran kebijakan itu sendiri. Dimana pelaksana kebijakan melakukan suatu aktivitas atau kegiatan. [25]

## **2.3 Preprocessing**

Preprocessing adalah proses pengolahan data asli yang akan dipersiapkan untuk diolah pada tahap selanjutnya. Masukan data awal yaitu berupa dokumen. Tahapan preprocessing yang dilakukan pada penelitian ini meliputi Tokenizing , proses Filtering, proses Case Folding, proses Stopword Removal, proses Stemming dan perhitungan bobot TF-IDF serta Normalisasi [7].

### **2.3.1 Tokenizing**

Proses Tokenizing adalah proses pemotongan string input berdasarkan tiap kata yang menyusunnya, memecahkan dokumen dan memisahkannya menjadi setiap kata. Pemecahan dokumen menjadi kata-kata tunggal dilakukan dengan menscan dokumen dan setiap kata teridentifikasi atau terpisahkan dengan kata yang lain oleh pemisah spasi [9]. Fitur ini terdiri dari tipe kapitalisasi (tipe case), keberadaan digit, tanda baca, karakter spesial dan lain sebagainya. Hasil keluaran dari proses tokenizing akan dipergunakan sebagai masukan dalam tahap transformasi teks. Tahapan-tahapan yang dilakukan dalam tokenizing yaitu :

1. Membaca data masukan berupa teks
2. Memeriksa apakah sudah EOF. Jika sudah maka algoritma berakhir, jika belum maka ke tahap 3.
3. Memeriksa teks dengan memecahnya menjadi kata (term). Mengecek apakah menemukan spasi. Jika bertemu spasi maka ke tahap 4. Jika tidak maka scan kembali.
4. Mengambil kata (term) dari teks dan membuat indeks baru untuk menandai asal kata (term).
5. Algoritma selesai.

### **2.3.2 Filtering**

Filtering mengacu pada proses memutuskan istilah mana yang harus digunakan untuk merepresentasikan dokumen sehingga dapat digunakan untuk menggambarkan isi dokumen dan membedakan dokumen dengan dokumen lain yang ada pada koleksi [10]. Dengan kata lain filtering merupakan proses dimana

teks selain karakter “a” sampai “z” dan spasi akan dihilangkan dan hanya menerima spasi. Tahapan-tahapan yang dilakukan yaitu :

1. Membaca data masukan berupa teks
2. Mengecek apakah ada huruf selain “a” sampai “z” dan “ ” (spasi). Jika tidak ditemukan, maka algoritma berakhir. Jika ada.
3. Hapus karakter yang bukan termasuk “a” sampai “z” dan “ ” (Spasi).
4. Algoritma selesai.

### **2.3.3 Case Folding**

Case folding adalah tahapan pemrosesan teks dimana semua teks diubah ke dalam case yang sama [11]. Pada penelitian ini semua huruf dalam teks dokumen diseragamkan menjadi huruf kecil . Tahapan-tahapan yang dilakukan yaitu :

1. Membaca data masukan berupa teks
2. Mengecek apakah ada huruf kapital (Jika menggunakan lower case). Jika ada maka ke tahap
3. jika tidak maka algoritma selesai. 3. Lakukan perubahan string lower pada setiap string.
4. Algoritma selesai

### **2.3.4 Stopword Removal**

Proses pembuangan stopwords dimaksudkan untuk mengetahui suatu kata masuk ke dalam stopwords atau tidak. Stoplist berisi kumpulan kata yang 'tidak relevan', tetapi seringkali muncul dalam sebuah dokumen. Dengan kata lain, stoplist berisi sekumpulan stopwords. Stopwords removal adalah proses menghilangkan kata yang 'tidak relevan' dari sebuah dokumen teks dengan cara membandingkannya dengan stoplist yang ada [12].

Stopword biasanya berisi kata-kata yang sering kali muncul berupa kata sambung, kata depan, kata ganti, kata penghubung, dll. Namun artinya tidak deskriptif dan tidak memiliki keterkaitan dengan topik tertentu. Tahapan-tahapan yang dilakukan yaitu :

1. Membaca data masukan berupa teks
2. Mengecek apakah ada kata yang termasuk Stopword Removal yang ada dalam kamus, jika tidak ada maka algoritma selesai, jika ada

3. Menghapus semua kata yang termasuk stopwords removal
4. Algoritma selesai.

### 2.3.5 Stemming

Stemming merupakan suatu proses yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (root word) dengan menggunakan aturan-aturan tertentu. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke root wordnya yaitu “sama”. Proses stemming pada teks berbahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan [13]. Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi:

***Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1***

Pada penelitian ini Algoritma yang digunakan dalam proses ini yaitu nazief dan andriani. Algoritma Stemming Bahasa Indonesia M. Adriani dan B Nazief ini mempunyai aturan imbuhan sendiri dengan model, seperti :

**[[[AW+]AW+]AW+] Kata-Dasar [[+AK][+KK][+P]]**

AW : Awalan

AK : Akhiran

KK : Kata Ganti kepunyaan

P : Partikel

Tanda kurung besar menandakan bahwa imbuhan adalah opsional [14]. Algoritma yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut:

1. Cari kata yang akan distem dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah root word. Maka algoritma berhenti.
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “- nya”), jika ada.

3. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a
  - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
  - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
  - a. Periksa tabel kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.
  - b. For  $i = 1$  to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.
5. Melakukan Recording.
6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai root word. Proses selesai.

Tipe awalan ditentukan melalui langkah-langkah berikut:

1. Jika awalnya adalah: “di-”, “ke-”, atau “se-” maka tipe awalnya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
2. Jika awalnya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalnya.
3. Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.
4. Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan dapat dilihat pada Tabel 2.2 Hapus awalan jika ditemukan.

**Tabel 2.2.1** Kombinasi Awalan Yang Tidak Diajukan

Awalan	Akhiran yang tidak diijinkan
be-	-i
di-	-an

ke-	-i, -kan
me-	-an
se-	-i, -kan

**Tabel 2.2.2** Cara Menentukan Awalan Untuk Kata Yang Diawali Dengan “te-”

Following Characters				Tipe Awalan
Set 1	Set 2	Set 3	Set 4	
“-r-“	“-r-“	-	-	none
“-r-“	vowel	-	-	ter-luluh
“-r-“	not (vowel or “-r-“)	“-er-“	vowel	ter
“-r-“	not (vowel or “-r-“)	“-er-“	not vowel	ter-
“-r-“	not (vowel or “-r-“)	not “-er-“	-	ter
not (vowel or “-r-“)	“-er-“	vowel	-	none
not (vowel or “-r-“)	“-er-“	not vowel	-	te

**Tabel 2.2.3** Jenis Awalan Berdasarkan Tipe Awalannya

Tipe Awalan	Awalan yang harus dihapus
di-	di-
ke-	ke-
se-	se-
te-	te-
ter-	ter-
ter-luluh	ter

### 2.3.6 Metode Pembobotan Kata TF-IDF

Term Frequency – Inverse Document Frequency digunakan untuk menentukan nilai frekuensi sebuah kata di dalam banyaknya dokumen. Perhitungan statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya dan seberapa relevannya sebuah kata di dalam sebuah dokumen[15].

Pembobotan diperoleh dari frekuensi jumlah kemunculan sebuah kata yang terdapat di dalam sebuah dokumen, term frequency (tf). Sebuah kata atau jumlah kemunculan term di dalam koleksi dokumen, inverse document frequency (idf). TF-IDF dapat berhasil digunakan dalam penyaringan di berbagai bidang, termasuk text summarization dan klasifikasi[15].

Bobot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu dokumen dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen [15].

Nilai idf sebuah term (kata) dapat dihitung menggunakan persamaan (2.1) berikut:

$$IDF_t = \log(N/df) \quad ( 2.1 )$$

Untuk menghitung bobot (W) masing-masing dokumen terhadap setiap term (kata) dapat menggunakan persamaan (2.2) berikut:

$$W_{dt} = tf_{dt} * IDF_t \quad ( 2.2 )$$

Dimana:

W = bobot dokumen ke-d terhadap kata ke-t

d = dokumen ke-d

t = kata ke-t

tf = banyaknya kata yang dicari pada sebuah dokumen

N = total dokumen

df = banyak dokumen yang mengandung tiap kata

Metode ini mampu menghitung bobot setiap term yang ada pada dokumen. Namun dalam metode ini tidak ada solusi untuk mengatasi resiko adanya redundansi (kemiripan) term yang dihasilkan dalam dokumen.

### 2.3.7 Normalisasi Bobot

Normalisasi dilakukan terhadap vektor fitur dokumen untuk menghilangkan pengaruh anggapan bahwa dokumen panjang lebih relevan dibandingkan dokumen pendek. Dengan normalisasi ini dapat membantu menormalkan batas nilai dengan melakukan standarisasi nilai ke dalam interval 0 sampai dengan 1. Sehingga dapat dituliskan sebagai berikut :

$$W_{ij} = \frac{tf_{ij} (\log(\frac{N}{n})+1)}{\sqrt{\sum_{k=1}^r (tf_{ij})^2 \times (\log(\frac{N}{n})+1)^2}} \quad ( 2.3 )$$

## 2.4 Seleksi Fitur

Tahap seleksi fitur bertujuan untuk mengurangi fitur yang redundan dan tidak relevan. Fitur yang tidak relevan adalah fitur yang tidak memberikan informasi berguna tentang data. Fitur redundan adalah fitur yang tidak memberikan informasi lebih banyak daripada fitur yang saat ini dipilih [16].

Algoritma seleksi fitur terbagi ke dalam tiga pendekatan yaitu pendekatan filter, pendekatan wrapper, dan pendekatan embedded [16]. Pendekatan Filter

mengevaluasi kualitas dari fitur yang diseleksi secara independen dari algoritma klasifikasi. Pendekatan Wrapper membutuhkan penerapan dari algoritma klasifikasi untuk mengevaluasi kualitas klasifikasi. Pendekatan Embedded menerapkan seleksi fitur selama pembelajaran dari parameter-parameter yang optimal.

Metode filter menyeleksi fitur yang relevan sebelum berpindah pada fase pembelajaran selanjutnya. Fitur yang terlihat paling signifikan dipilih untuk klasifikasi, sementara sisanya yang lain disisihkan. Salah satu metode yang digunakan untuk melakukan Filter adalah Information Gain [18].

#### 2.4.1 Information Gain

Information Gain adalah salah satu algoritma seleksi fitur yang digunakan untuk memilih fitur terbaik. Nilai Information Gain yang didapat akan digunakan untuk menyeleksi fitur menggunakan threshold sehingga menghasilkan fitur terbaik [5]. Nilai Information Gain dapat didefinisikan dengan persamaan (2.4) sebagai berikut [16] :

$$IG(t) = -\frac{A+C}{N} \log\left(\frac{A+C}{N}\right) + \frac{A}{N} \log\left(\frac{A}{A+B}\right) + \frac{C}{N} \log\left(\frac{C}{C+D}\right) \quad (2.4)$$

Dimana :

A : jumlah dokumen dikelas k yang mengandung term t

B : jumlah dokumen diluar kelas k yang mengandung term t

C : jumlah dokumen dikelas k yang tidak mengandung term t

D : jumlah dokumen diluar kelas k yang tidak mengandung term t

N : Banyak dokumen

Setelah nilai Information Gain didapat, maka dilakukan langkah selanjutnya adalah menentukan nilai threshold. Nilai Information Gain yang paling kecil dapat dijadikan sebagai nilai threshold dengan akurasi terbaik [17]. Apabila nilai Information Gain threshold bernilai kecil, mengakibatkan informasi penting terlalu banyak sehingga dapat menimbulkan noise karena masih luasnya lingkup fitur penting [5].

## 2.5 Metode Klasifikasi Support Vector Machine (SVM)

SVM dikembangkan oleh Boser, Guyon, Vapnik pada tahun 1992. SVM merupakan suatu teknik untuk menemukan hyperplane yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). Hyperplane adalah garis batas pemisah data antar kelas, sedangkan margin adalah jarak antara hyperplane dengan data terdekat pada masing-masing kelas. Adapun data terdekat dengan hyperplane pada masing-masing kelas inilah yang disebut support vector. Karena konsep awal SVM adalah untuk mengatasi masalah klasifikasi dua kelas maka, ada dua metode diusulkan agar SVM bisa digunakan untuk klasifikasi multi-class dengan pendekatan mengkombinasikan beberapa binary classifier atau menggabungkan semua data yang terdiri dari beberapa kelas ke dalam permasalahan optimasi[19].

Dalam linear Support Vector Machine pemisah merupakan fungsi linear. Data latih dinyatakan oleh  $(x_i, y_i)$  dan  $x_i = \{x_1, x_2, \dots, x_q\}$  merupakan atribut (fitur) set untuk data latih kelas ke- $i$ . Untuk  $y_i \in \{-1, 1\}$  menyatakan label kelas. pendefinisian persamaan suatu hyperplane pemisah yang dituliskan dengan:

$$w * x_i + b = 0 \quad (2.5)$$

Data  $x_i$  yang terbagi ke dalam dua kelas, yang termasuk kelas -1 (sampel negatif) didefinisikan sebagai vektor yang memenuhi pertidaksamaan (2.6) berikut.

$$w * x_i + b > 0 \text{ untuk } y_i = -1 \quad (2.6)$$

Sedangkan yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan (2.7) berikut.

$$w * x_i + b > 0 \text{ untuk } y_i = +1 \quad (2.7)$$

Dimana:

$x_i$  = data input

$y_i$  = label yang diberikan

$w$  = nilai dari bidang normal

$b$  = posisi bidang relatif terhadap pusat koordinat

Parameter  $w$  dan  $b$  adalah parameter yang akan dicari nilainya. Bila label data  $y_i = -1$ , maka pembatas menjadi persamaan (2.8) berikut:

$$w * x_i + b \leq -1 \quad (2.8)$$

Bila label data  $y_i = +1$ , maka pembatas menjadi persamaan (2.9) berikut:

$$w * x_i + b \geq +1 \quad (2.9)$$

Margin terbesar dapat dicari dengan cara memaksimalkan jarak antar bidang pembatas kedua kelas dan titik terdekatnya, yaitu  $2/|w|$ . Hal ini dirumuskan sebagai permasalahan quadratic programming (QP) problem yaitu mencari titik minimal persamaan (2.10) dengan memperhatikan persamaan (2.11) berikut:

$$\min r(w) = \frac{1}{2} \|w\|^2 \quad (2.10)$$

$$y_i(w * x_i + b) - 1 \geq 0, (i = 1, \dots, n) \quad (2.11)$$

Permasalahan ini dapat dipecahkan dengan membagi teknik komputasi. Lebih mudah diselesaikan dengan mengubah persamaan (2.10) ke dalam fungsi Lagrangian pada persamaan (2.12), dan menyederhanakannya menjadi persamaan (2.13) berikut:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i (y_i((w^T x_i + b) - 1)) \quad (2.12)$$

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T x_i + b) + \sum_{i=1}^n a_i \quad (2.13)$$

Dimana  $a_i$  adalah lagrange multiplier yang bernilai nol atau positif ( $a_i \geq 0$ ). Nilai optimal dari persamaan (2.13) dapat dihitung dengan meminimalkan  $L$  terhadap  $w$ ,  $b$  dan  $a$ . Dapat dilihat pada persamaan (2.14) sampai (2.16) berikut:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n a_i y_i x_i = 0 \quad (2.14)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n a_i y_i = 0 \quad (2.15)$$

$$\frac{\partial L}{\partial a} = \sum_{i=1}^n a_i y_i (w^T x_i + b) - \sum_{i=1}^n a_i = 0 \quad (2.16)$$

Maka masalah Lagrange untuk klasifikasi dapat dinyatakan pada persamaan (2.17) berikut:

$$\text{Min } L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w^T x_i + b) - \sum_{i=1}^n a_i \quad (2.17)$$

Dengan memperhatikan persamaan (2.18) dan (2.19) berikut:

$$w - \sum_{i=1}^n a_i y_i x_i = 0 \quad (2.18)$$

$$\sum_{i=1}^n a_i y_i = 0 \quad (2.19)$$

Model persamaan (2.17) diatas merupakan model primal Lagrange. Sedangkan dengan memaksimalkan L terhadap  $a_i$  , persamaannya menjadi persamaan (2.20) berikut:

$$\text{Max} \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1, j=1}^n a_i a_j y_i y_j^T x_i x_j^T \quad ( 2.20 )$$

Dengan memperhatikan persamaan (2.21) berikut:

$$\sum_{i=1}^n a_i y_i = 0, a_i \geq 0 (i, j = 1, \dots, n) \quad ( 2.21 )$$

Untuk mencari nilai  $x_i$  dan  $y_i$  dapat dilakukan ketika sudah didapatkan nilai tiap kata (term) dari pembobotan tf-idf dan inisialisasi kelas . Hasil dari pembobotan tfidf diubah ke dalam bentuk format data svm, sedangkan data kelas menjadi label data svm.

Untuk mendapatkan nilai  $a_i$  , langkah pertama adalah mengubah setiap abstrak menjadi nilai vektor (support vector) =  $( x \ y )$  . Kemudian nilai vektor dari setiap abstrak dimasukkan ke persamaan (2.22) kernel trick phi berikut.

$$\phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{cases} \sqrt{x_n^2 + y_n^2} > 2 \text{ maka } \begin{bmatrix} \sqrt{x_n^2 + y_n^2} - x + |x - y| \\ \sqrt{x_n^2 + y_n^2} - y + |x - y| \end{bmatrix} \\ \sqrt{x_n^2 + y_n^2} \leq 2 \text{ maka } \begin{bmatrix} x \\ y \end{bmatrix} \end{cases} \quad ( 2.22 )$$

Nilai x didapatkan dari persamaan (2.23) kernel linear untuk x berikut:

$$\sum_{i=1, j=1}^n x_i x_j^T, (i, j = 1, \dots, n) \quad ( 2.23 )$$

Nilai y didapatkan dari persamaan (2.24) kernel linear untuk y berikut:

$$\sum_{i=1, j=1}^n y_i y_j^T, (i, j = 1, \dots, n) \quad ( 2.24 )$$

Untuk mendapatkan jarak tegak lurus yang optimal dengan mempertimbangkan vektor positif, maka hasil perhitungan dari substitusi nilai x dan nilai y ke persamaan (2.22) diberi nilai bias = 1 . Kemudian cari parameter  $a_i$  , dengan terlebih dahulu mencari nilai fungsi setiap abstrak menggunakan persamaan (2.25), lalu mencari nilai  $a_i$  pada persamaan linear menggunakan persamaan (2.26) dengan memperhatikan  $i, j = 1, \dots, n$  berikut:

$$\sum_{i=1, j=1}^n x_i S_i^T S_j \quad ( 2.25 )$$

$$\sum_{i=1, j=1}^n x_i S_i^T S_j = y_i \quad ( 2.26 )$$

Setelah parameter  $a_i$  didapatkan, kemudian masukkan ke persamaan (2.27) berikut:

$$\tilde{W} = \sum_{i=1}^n a_i S_i \quad ( 2.27 )$$

Hasil yang didapatkan menggunakan persamaan (2.27), selanjutnya digunakan persamaan (2.28) untuk mendapatkan nilai  $w$  dan  $b$ :

$$y = wx + b \quad ( 2.28 )$$

Sedemikian sehingga didapatkanlah nilai  $w$  dan nilai  $b$  atau nilai hyperplane untuk mengklasifikasikan kedua kelas.

Sebuah fungsi bisa menjadi fungsi kernel jika memenuhi Teorema Mercer, yang menyatakan bahwa matriks kernel yang dihasilkan harus bersifat semi positive semi definite[20]. berikut ini adalah beberapa fungsi kernel yang umum digunakan yaitu:

a. Kernel linier

$$K(x_i, x) = x_i^T x$$

b. Polynomial

$$K(x_i, x) = (\gamma \cdot x_i^T x + r)^p, \gamma > 0$$

c. Radial basis function

$$K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0$$

d. Sigmoid kernel

$$K(x_i, x) = \tanh(\gamma x_i^T x + r)$$

## 2.6 Smooth Support Vector Machine

SSVM adalah pengembangan SVM dengan menggunakan teknik smoothing. Metode ini pertama kali diperkenalkan oleh Lee pada tahun 2001. SVM memanfaatkan optimasi dengan quadratic programming, sehingga untuk data berdimensi tinggi dan data jumlah besar SVM menjadi kurang efisien. Oleh karena itu dikembangkan smoothing technique yang menggantikan plus function SVM

dengan integral dari fungsi sigmoid neural network yang selanjutnya dikenal dengan Smooth Support Vector Machine (SSVM) [21].

Diberikan masalah klasifikasi dari  $n$  objek dalam ruang dimensi  $R^p$  sehingga susunan data berupa matriks  $A$  berukuran  $n \times p$  dan keanggotaan tiap titik terhadap kelas  $\{+1\}$  atau  $\{-1\}$  yang didefinisikan pada diagonal matriks  $D$  berukuran  $n \times n$ , problem optimasinya adalah

$$\min_{w,y} \frac{c}{2} y'y + \frac{1}{2} (w'w + b^2) \quad (2.29)$$

dengan kendala

$$D(Aw + eb) + y \geq e, y \geq 0. \quad (2.30)$$

Solusi problem adalah

$$y = (e - D(Aw + eb)) \quad (2.31)$$

Dimana  $y$  adalah variabel slack yang mengukur kesalahan klasifikasi. Kemudian dilakukan substitusi dan konversi, sehingga persamaan dapat ditulis sebagai berikut:

$$\min_{w,y} \frac{c}{2} \| (e - D(Aw - eb)) \|_2^2 + \frac{1}{2} (w'w + b^2) \quad (2.32)$$

Fungsi objektif dalam persamaan tidak memiliki turunan kedua. Teknik smoothing yang diusulkan dilakukan dengan mengganti fungsi plus dengan  $p(x,a)$  yaitu integral dari fungsi sigmoid neural network atau dapat dituliskan sebagai berikut:

$$p(x,a) = x + \frac{1}{a} \log(1 + \exp(-ax)), a > 0 \quad (2.33)$$

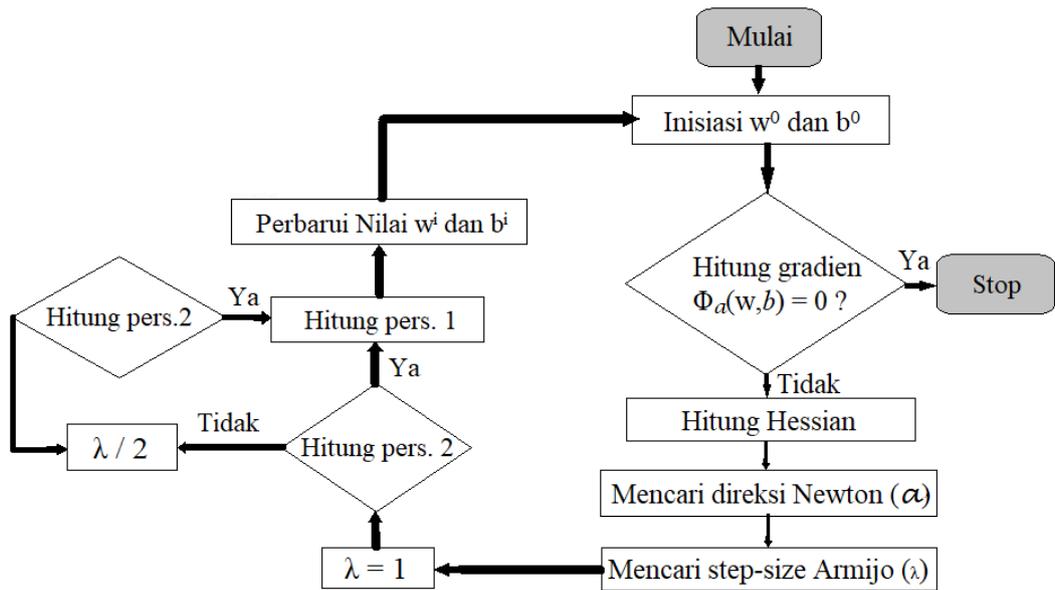
dimana  $a$  adalah parameter smoothing. Dengan menggantikan fungsi plus dengan  $p(x,a)$  maka diperoleh model SSVM sebagai berikut:

$$\min_{w,b} \frac{c}{2} \| (e - D(Aw - eb)) \|_2^2 + \frac{1}{2} (w'w + b^2) \quad (2.34)$$

Secara umum, problem optimasi SSVM dapat ditulis sebagai berikut:

$$\min_{(w,y) \in R^{(w,y)}} = \min_{(w,y) \in R^{(n+1)}} \frac{c}{2} \| p(e - D(Aw - ey), a) \|_2^2 + \frac{1}{2} (w'w + y^2) \quad (2.35)$$

Yang diselesaikan dengan iterasi Newton Armijo (Gambar 2.1) dan  $K(X_i, X_j)$  merupakan fungsi kernel yang dalam penelitian ini digunakan kernel Gaussian atau bisa dirumuskan berikut  $K(X_i, X_j) = \exp(-\gamma(\|X_i - X_j\|_2)^2)$  dengan parameter kernel  $\gamma$ .



**Gambar 2.1.1 Diagram Alir Algoritma Newton-Armijo**

Persamaan 1 :

$$\phi_a(w_i, b_i) - \phi_a((w_i, b_i) + (\lambda_i d_i)) \geq -\delta \lambda_i \nabla \phi_a(w_i, b_i) d_i \quad (2.36)$$

Persamaan 2 :

$$w_{i+1}, b_{i+1} = (w_i, b_i) + (\lambda_i d_i) \quad (2.37)$$

Saat iterasi pada algoritma Newton-Armijo berhenti, diperoleh nilai  $w$  dan  $b$  yang konvergen. Dengan demikian fungsi pemisah yang diperoleh untuk kasus klasifikasi linier adalah

$$F(x) = \text{sign}(w'x + b) \quad (2.38)$$

Sedangkan fungsi pemisah untuk kasus klasifikasi nonlinier adalah sebagai berikut :

$$F(x) = \text{sign}(w'x + b) = \text{sign}(u'D'K(X_i, X_j) + b) \quad (2.39)$$

Perumusan program linier SVM 1-norm adalah salah satu cara untuk memilih atribut (feature selection) di antara varian-varian norm SVM, problem linier tersebut adalah sebagai berikut:

$$\min_{(w,b,s,\xi) \in \mathbb{R}^{(2p)+1+n}} Ce' \xi + e's \quad (2.40)$$

Dengan kendala

$$D(Aw + eb) + y \geq e \quad (2.41)$$

$$\begin{aligned} -s \leq w \leq s \\ y \geq 0 \end{aligned} \quad ( 2.41 )$$

Solusi dari  $w$  mampu menghasilkan model yang parsimoni dan bersifat sparsity. Jika nilai dari elemen vektor  $w_p = 0$ , maka variabel  $p$  tidak berkontribusi dalam penentuan kelas. Kontribusi atribut atau variabel prediktor dapat dinilai dari besarnya nilai  $w_i$  untuk masing-masing atribut, dengan  $i=1,2, \dots, p$ . [21]

Support Vector Machine (SVM) adalah suatu sistem pembelajaran yang menggunakan ruang hipotesis dari suatu fungsi linear dalam suatu ruang dimensi berfitur tinggi yang dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory.

## 2.7 Perangkat Lunak

### 2.7.1 Python

Python adalah bahasa pemrograman komputer, sama seperti bahasa pemrograman lain, misalnya C, C++, Pascal, Java, PHP, Perl dan lain - lain. Sebagai bahasa pemrograman, Python tentu memiliki dialek, kosakata atau kata kunci, dan aturan tersendiri yang jelas berbeda dengan bahasa pemrograman lainnya [3].

Bahasa pemrograman Python disusun di akhir tahun 1980-an dan implementasinya baru dimulai pada Desember 1989 oleh Guido Van Rossum di Centrum Wiskunde & Information (CWI), sebuah pusat riset di bidang matematika dan sains, Amsterdam – Belanda; sebagai suksesor atau pengganti dari bahasa pemrograman pendahulunya, bahasa pemrograman ABC, yang juga dikembangkan di CWI oleh Leo Geurts, Lambert Meertens, dan Steven Pemberton [3]. Secara umum, para programmer banyak yang menjatuhkan pilihannya ke bahasa Python karena alasan – alasan berikut :

1. Python memiliki konsep design yang bagus dan sederhana, yang berfokus pada kemudahan dalam penggunaan. Kode Python dirancang untuk mudah dibaca, dipelajari, digunakan ulang, dan dirawat. Selain itu, Python juga mendukung pemrograman berorientasi objek dan pemrograman fungsional.

2. Python dapat meningkatkan produktivitas dan menghemat waktu bagi para programmer. Untuk memperoleh hasil program yang sama, kode Python jauh lebih sedikit dibandingkan dengan kode yang ditulis menggunakan bahasa – bahasa pemrograman lain.
3. Program yang ditulis menggunakan Python dapat dijalankan di hampir semua sistem operasi (Linux, Windows, Mac, dan lain – lain), termasuk untuk perangkat – perangkat mobile
4. Python bersifat gratis atau bebas (free) dan open-source, meskipun digunakan untuk kepentingan komersil.

### **2.7.2 Python 2.7**

Sejak Python 2.6, komunitas Python menyertakan modul `_future_` ke dalam Python 2. Dengan menggunakan modul ini, para pengguna Python 2 masih dapat menikmati fitur – fitur yang ada terdapat di dalam Python 3. Python 2 dilengkapi dengan tool bernama 2to3, yang digunakan untuk mengkonversi kode Python 2 ke Python 3 [6].

### **2.7.3 Python 3**

Python 3 dikembangkan dengan tujuan jangka panjang. Fasilitas – fasilitas baru yang saat ini ditambahkan ke dalam bahasa Python hanya diimplementasikan ke dalam Python 3 dengan harapan para programmer yang masih menggunakan Python 2 secara bertahap bisa beralih ke Python 3 [6].

## **2.8 UML (Unified Modeling Language)**

UML (Unified Modeling Language) adalah notasi yang lengkap untuk membuat visualisasi model suatu sistem. Sistem berisi informasi dan fungsi, tetapi secara normal digunakan untuk memodelkan sistem komputer. Di dalam pemodelan objek guna menyajikan sistem yang berorientasi objek kepada orang lain, akan sangat sulit dilakukan dalam bentuk kode bahasa pemrograman [22].

UML disebut sebagai bahasa pemodelan bukan metode. Bahasa pemodelan (sebagian besar grafik) merupakan notasi model yang digunakan untuk mendesain secara cepat. Bahasa pemodelan merupakan bagian terpenting dari metode. UML

merupakan bahasa standar untuk penulisan blueprint software yang digunakan untuk visualisasi, spesifikasi, pembentukan dan pendokumentasian

+n alat-alat dari sistem perangkat lunak. UML biasanya disajikan dalam diagram atau gambar yang meliputi class beserta atribut dan operasinya, serta hubungan antar kelas. UML terdiri dari banyak diagram diantaranya use case diagram, activity diagram, class diagram dan sequence diagram.

### **2.8.1 Use Case Diagram**

Dalam konteks UML, tahap konseptualisasi dilakukan dengan pembuatan use case diagram yang sesungguhnya merupakan deskripsi peringkat tinggi bagaimana perangkat lunak (aplikasi) akan digunakan oleh penggunanya.

Use case diagram merupakan deskripsi lengkap tentang interaksi yang terjadi antara para actor dengan sistem[23]. Dalam hal ini, setiap objek yang berinteraksi dengan sistem merupakan actor untuk sistem, sementara use case merupakan deskripsi lengkap tentang bagaimana sistem berperilaku kepada actornya. Actor dalam use case diagram digambarkan sebagai ikon yang berbentuk manusia, sementara use case digambarkan sebagai elips yang berisi nama use case yang bersangkutan. Untuk mempermudah pemahaman, actor biasanya dituliskan sebagai kata benda, sementara use case biasanya dituliskan sebagai kata kerja.

### **2.8.2 Activity Diagram**

Activity diagram pada dasarnya menggambarkan skenario secara grafis. Serta activity diagram cukup serupa dengan diagram alir (flow chart), yang membedakan mungkin hanya swimlane yang menunjukkan suatu state berada pada objek/kelas tertentu. Keunggulan dari activity diagram adalah bahwa diagram tersebut lebih mudah dipahami dibanding skenario. Selain itu, dengan menggunakan activity diagram, kita dapat melihat dibagian manakah sistem dari suatu skenario akan berjalan[23].

### **2.8.3 Class Diagram**

Diagram kelas atau class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Class diagram menggambarkan struktur dan deskripsi class, package dan objek beserta hubungan satu sama lain seperti containment, pewarisan, asosiasi, dan lain-lain [7].

#### 2.8.4 Sequence Diagram

Diagram sekuen menggambarkan kelakuan/perilaku objek pada use case dengan mendeskripsikan waktu hidup objek dan message yang dikirimkan dan diterima antar objek. Objek-objek yang berkaitan dengan proses berjalannya operasi diurutkan dari kiri ke kanan berdasarkan waktu terjadinya dalam pesan yang terurut. Oleh karena itu untuk menggambar diagram sekuen maka harus diketahui objek-objek yang terlibat dalam sebuah use case beserta metode-metode yang dimiliki kelas yang diinstansiasi menjadi objek itu [7].

#### 2.9 Akurasi

Untuk menghitung akurasi dari SVM dilakukan perhitungan dengan rumus [24] :

$$Akurasi = \frac{Jumlah\ diklasifikasi\ secara\ benar}{total\ sample\ testing\ yang\ diuji} \quad ( 2.41 )$$