



TESIS - PM 147501

**PREDIKSI KINERJA MAHASISWA MENGGUNAKAN  
*SUPPORT VECTOR MACHINE* UNTUK  
PENGELOLA PROGRAM STUDI DI PERGURUAN  
TINGGI (Studi Kasus: Program Studi Magister  
Statistika ITS)**

**FATHIN HILMIYAH**  
9115 205 311

**DOSEN PEMBIMBING**  
Prof. Ir. Joko Lianto Buliali, M.Sc, Ph.D.

**DEPARTEMEN MANAJEMEN TEKNOLOGI  
BIDANG KEAHLIAN MANAJEMEN TEKNOLOGI INFORMASI  
FAKULTAS BISNIS DAN MANAJEMEN TEKNOLOGI  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017**

# LEMBAR PENGESAHAN

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Manajemen Teknologi (M.MT)  
di  
Institut Teknologi Sepuluh Nopember

Oleh :

**FATHIN HILMIYAH**  
NRP. 9115205311

Tanggal Ujian : 11 Juli 2017  
Periode Wisuda : September 2017

Disetujui oleh :

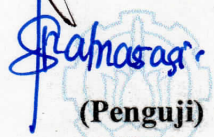
1. **Prof. Ir. Joko Lianto Buliali, M.Sc, Ph.D**  
NIP. 19670727 199203 1 002

  
(Pembimbing)

2. **Dr. Tech. Ir. R.V. Hari Ginardi, M.Sc**  
NIP. 19650518 199203 1003

  
(Penguji)

3. **Dr. Vita Ratnasari, S.Si, M.Si**  
NIP. 19700910 199702 2001

  
(Penguji)

Dekan Fakultas Bisnis dan Manajemen Teknologi

  
  
**Prof. Dr. Ir. Udisubakti Ciptomulyono, M. Eng.Sc**  
NIP. 19590318 198701 1 001

**PREDIKSI KINERJA MAHASISWA MENGGUNAKAN *SUPPORT VECTOR MACHINE* UNTUK PENGELOLA PROGRAM STUDI DI PERGURUAN TINGGI (Studi Kasus: Program Studi Magister Statistika ITS)**

Nama : Fathin Hilmiyah  
NRP : 9115205311  
Dosen Pembimbing : Prof. Ir. Joko Lianto Buliali, M.Sc, Ph.D

**ABSTRAK**

Kesuksesan sebuah perguruan tinggi sering dilihat dari keberhasilan prestasi belajar mahasiswa yang bernaung dalam institusi tersebut. Salah satu langkah yang dibutuhkan yakni dengan melakukan prediksi kinerja akademik bagi calon mahasiswa. Prediksi kinerja mahasiswa ini dapat digunakan sebagai salah satu penunjang keputusan evaluasi kinerja mahasiswa. Program Studi Magister Statistika ITS Surabaya belum memiliki *tools* untuk memprediksi kinerja dan keberhasilan tersebut. Maka dari itu, peneliti melakukan penelitian untuk membuat pemodelan Prediksi Kinerja Mahasiswa untuk Pengelola Program Studi di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika ITS). Penelitian ini menggunakan data akademik berupa sampel data mahasiswa Magister Statistika ITS sejumlah 318 data. Metode yang digunakan adalah metode *Support Vector Machine* (SVM). Pelatihan dan pengujian sistem dilakukan dengan metode *10-fold Cross Validation* dengan mengukur hasil akurasi, presisi dan *recall*. Berdasarkan uji performa perbandingan fungsi Kernel, *Linear* Kernel merupakan fungsi yang paling cocok untuk menghasilkan prediksi paling optimal. Dari uji coba yang dilakukan, penerapan SVM memiliki hasil perhitungan akurasi, presisi dan *recall* yang lebih baik bila dibandingkan dengan penggunaan metode Regresi Logistik. Berdasarkan uji pengaruh setiap variabel, performansi kinerja mahasiswa dapat ditingkatkan secara signifikan dengan mempertimbangkan beberapa indikator mahasiswa yakni asal daerah, status kerja, jalur masuk, nilai toefl, nilai TPA, akreditasi asal institusi, IPK, lama studi dan waktu tunggu mahasiswa.

**Kata Kunci:** Prediksi, Mahasiswa, *Support Vector Machine*

*halaman ini sengaja dikosongkan*

**PREDICTING STUDENT PERFORMANCE USING SUPPORT VECTOR  
MACHINE FOR MANAGEMENT OF HIGHER EDUCATION  
(Case Study: Magister Study Program, Department of Statistics ITS)**

Name : Fathin Hilmiyah  
NRP : 9115205311  
Supervisor : Prof. Ir. Joko Lianto Buliali, M.Sc, Ph.D

**ABSTRACT**

*The success of any higher educational institutions is often seen from the success of their student achievement. One of the steps needed is to predict academic performance for students. These student performance predictions can be used to improve the quality of managerial decisions and to impart quality education. Magister Study Program at Department of Statistics ITS Surabaya does not have the tools to predict the performance of their students yet. Therefore, researchers conduct this paper to make modeling of Prediction Student Performance for Management of Higher Education at Magister Study Program, Department of Statistics ITS using Support Vector Machine (SVM). The dataset comprises of 318 student records that provided information about student demographics and previous academic standings in which 12 significant variable of students were extracted for experimentation in the study. Training and system testing is done by 10-fold Cross Validation method by measuring accuracy, precision and recall. Based on comparative performance Kernel function for SVM, the result show that Linear Kernel is the most suitable Kernel function to produce the most optimal prediction for student performance. From the experiments performed, the application of SVM has a better accuracy, precision and recall calculation results when compared with the use of logistic regression method. Based on the test of the influence of each variable, the quality of student performance can be improved significantly by considering some student indicator they are work status, toefl score, TPA score, accreditation of institution before, GPA, study duration and waiting time of student from graduate to join magister program.*

**Kata Kunci:** *Prediction, Student Performance, Support Vector Machine*

*halaman ini sengaja dikosongkan*

## **KATA PENGANTAR**

Puji syukur kepada Allah SWT yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Tesis ini. Tesis ini disusun sebagai salah satu syarat untuk menyelesaikan studi Magister di Program Magister Manajemen Teknologi, Bidang Keahlian Manajemen Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Dengan ini, penulis menyampaikan penghormatan dan mengucapkan terima kasih kepada segala pihak yang telah memberikan bantuan berupa moral maupun material secara langsung maupun tidak langsung, kepada:

1. Dr. Ir. Mokhammad Suef, M.Sc.(Eng) selaku Kepala Departemen Manajemen Teknologi.
2. Dr. Tech. Ir. R. V. Hari Ginardi, M.Sc selaku dosen wali yang telah banyak memberikan arahan akademik kepada penulis.
3. Prof. Ir. Joko Lianto Buliali, M.Sc, Ph.D selaku dosen pembimbing tesis yang telah meluangkan waktu, tenaga dan pikiran dalam memberikan bimbingan, pengarahan dan ilmu pengetahuan kepada penulis.
4. Seluruh dosen pengajar yang telah memberikan bimbingan, pengajaran dan ilmu pengetahuan kepada penulis dan seluruh karyawan MMT-ITS Surabaya yang telah banyak membantu dalam berbagai hal selama masa perkuliahan.
5. Kedua orang tua H. M. Zuhdi Zen (Alm.) dan Dra. Nurul Latifah, Kakak, Muhammad Faruk A R dan Lailatul Habibah Ainy A R, serta seluruh keluarga besar yang telah memberikan dukungan, semangat, doa dan motivasi selama masa perkuliahan dan selama penulisan penelitian tesis ini.
6. Keluarga Besar Departemen Statistika ITS tempat penulis bekerja, yang telah memberikan izin melanjutkan studi dan tidak segan membantu dan memberi masukan keilmuan dalam berbagai hal selama masa perkuliahan.
7. Seluruh teman-teman MMT ITS angkatan 2015 yang telah memberikan motivasi dan masukan kepada penulis dalam menyusun tesis ini.

Akhir kata, penulis menyadari bahwa tesis ini masih jauh dari kesempurnaan dan memiliki banyak kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis mengharapkan kepada pembaca untuk dapat memberikan masukan yang dapat membangun untuk perbaikan di masa depan.

Surabaya, Juli 2017

Penulis



## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
ABSTRAK .....	iii
<i>ABSTRACT</i> .....	v
KATA PENGANTAR .....	vii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xiii
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian .....	4
1.4. Batasan Penelitian .....	4
1.5. Manfaat Penelitian .....	5
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>7</b>
2.1. Proses Klasifikasi .....	7
2.2. <i>Support Vector Machine</i> .....	9
2.3. Fungsi Kernel .....	14
2.4. Regresi Logistik sebagai Metode Pembanding .....	16
2.5. Evaluasi Performansi Metode Klasifikasi .....	17
2.6. <i>N-Folds Cross Validation</i> .....	18
2.7. Penelitian Terdahulu .....	18
2.8. Posisi Penelitian .....	32
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>37</b>
3.1. Jenis Penelitian .....	37
3.2. Diagram Alir Penelitian .....	38
3.3. Pengumpulan Data .....	42
3.4. Identifikasi Variabel .....	45

<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>51</b>
4.1. Distribusi Data .....	51
4.2. Implementasi SVM pada Matlab .....	54
4.3. Perhitungan Ketepatan Klasifikasi .....	58
4.4. Pelatihan dan Pengujian Model .....	61
4.5. Analisis Hasil Penerapan SVM .....	69
4.6. Model Prediksi Kinerja Mahasiswa menggunakan SVM .....	77
<b>BAB V PENUTUP .....</b>	<b>79</b>
5.1. Kesimpulan .....	79
5.2. Saran .....	80
<b>DAFTAR PUSTAKA .....</b>	<b>xv</b>
<b>BIOGRAFI PENULIS .....</b>	<b>xix</b>

## DAFTAR TABEL

Tabel 2.1.	Akurasi Klasifikasi .....	17
Tabel 2.2.	Variabel dengan Potensial Tinggi (Kaur, P. dkk., 2015) .....	19
Tabel 2.3.	Variabel untuk Prediksi (Asogwa, O.C. dan Oladugba, A.V, 2015) .....	22
Tabel 2.4.	Perbandingan Hasil Klasifikasi dengan Data Sebenarnya (Asogwa, O. C. dan Oladugba, A. V., 2015) .....	23
Tabel 2.5.	Karakteristik Mahasiswa untuk Prediksi Kegagalan Akademis (Teshnizi, S. H., dkk., 2015) .....	24
Tabel 2.6.	Deskripsi Atribut dari Data <i>Set</i> (Shaziya, H., dkk., 2015) .....	26
Tabel 2.7.	Hasil Pengujian <i>Support Vector Machine</i> (Nurhayati,S., dkk., 2015) ...	28
Tabel 2.8.	Tingkat Akurasi Klasifikasi SVM (Theopilus, B. S., dkk., 2014) .....	30
Tabel 2.9.	Performansi Fungsi Kernel dengan parameter $C=4096$ (Parikh, K. S. dan Shah, T. P., 2016) .....	31
Tabel 2.10.	Lingkup dan Metode yang Digunakan pada Penelitian Terdahulu .....	33
Tabel 3.1.	Variabel dan Definisi Operasional Variabel .....	44
Tabel 3.2.	Variabel Masukan Tanpa Proses Konversi .....	46
Tabel 3.3.	Variabel Masukan dengan Proses Konversi .....	47
Tabel 3.4.	Variabel Keluaran Level Kinerja Mahasiswa .....	48
Tabel 4.1.	Distribusi Data Variabel Masukan .....	51
Tabel 4.2.	Distribusi Data Variabel Keluaran .....	53
Tabel 4.3.	Nilai Optimal Parameter pada Fungsi Kernel .....	55
Tabel 4.4.	Konsep <i>One Against All SVM</i> untuk Klasifikasi Level Mahasiswa .....	57
Tabel 4.5.	Konsep <i>Confusion</i> Matriks untuk <i>Output</i> Multikelas .....	59
Tabel 4.6.	Hasil Prosentase Akurasi Uji Performa Fungsi Kernel .....	60
Tabel 4.7.	Hasil Prosentase Presisi Uji Performa Fungsi Kernel .....	61
Tabel 4.8.	Hasil Prosentase <i>Recall</i> Uji Performa Fungsi Kernel .....	62
Tabel 4.9.	<i>Confusion</i> Matriks Hasil Prediksi Kinerja Mahasiswa .....	69
Tabel 4.10.	<i>Confusion</i> Matriks Hasil Prediksi Kinerja Mahasiswa Optimasi FP ....	70
Tabel 4.11.	Kelompok Performansi Mahasiswa .....	71
Tabel 4.12.	Prediksi Kinerja dengan Pengelompokan Level Performansi .....	72

*halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

Gambar 2.1. Ilustrasi Klasifikasi Dua Kelas .....	8
Gambar 2.2. Ilustrasi Klasifikasi Multi Kelas .....	8
Gambar 2.3. Ilustrasi Mencari Fungsi Pemisah Optimal dalam Kasus Klasifikasi.....	11
Gambar 2.4. Fungsi $\Phi$ memetakan data ke ruang <i>vector</i> yang berdimensi lebih tinggi .....	12
Gambar 2.5. Komparasi Uji Akurasi Masing-masing Algoritma (Kaur, P. dkk., 2015) .....	20
Gambar 2.6. Model Proses Prediksi Ujian Akhir Semester Mahasiswa (Shaziya, H., dkk., 2015) .....	26
Gambar 2.7. Grafik Hasil Akurasi, F-Score dan G-Score untuk Macam-macam Kernel (Parikh, K. S. dan Shah, T. P., 2016) .....	32
Gambar 3.1. Kerangka Tahapan Penelitian .....	39
Gambar 3.2. Diagram Proses Pengolahan Data .....	40
Gambar 3.3. <i>Flowchart</i> Algoritma SVM .....	41
Gambar 4.1. Grafik Distribusi Data Variabel Masukan .....	52
Gambar 4.2. Grafik Distribusi Data Variabel Keluaran .....	54
Gambar 4.3. Konsep <i>One Against All</i> SVM Klasifikasi Level Mahasiswa .....	57
Gambar 4.4. Grafik Hasil Uji Coba Performansi Fungsi Kernel .....	63
Gambar 4.5. Grafik Hasil Akurasi dengan SVM dan Regresi Logistik .....	64
Gambar 4.6. Grafik Hasil Presisi dengan SVM dan Regresi Logistik .....	65
Gambar 4.7. Grafik Hasil <i>Recall</i> Metode SVM dan Regresi Logistik .....	66
Gambar 4.8. Grafik Hasil Uji Coba Performansi Metode Pembanding .....	67
Gambar 4.9. Grafik Hasil Prediksi Pengelompokan Performansi Kinerja .....	73

*halaman ini sengaja dikosongkan*

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Perguruan Tinggi merupakan institusi pendidikan akademik yang memiliki peran untuk mencetak sumber daya manusia berkualitas agar mampu memenuhi kualifikasi tuntutan pasar dunia kerja. Perguruan tinggi adalah tingkatan universitas yang terdiri atas sejumlah fakultas yang menyelenggarakan pendidikan akademik dalam sejumlah disiplin ilmu tertentu (Barthos,1992). Di Indonesia, jumlah perguruan tinggi yang semakin marak dengan berbagai macam disiplin ilmu mengakibatkan tingkat persaingan sesama perguruan tinggi juga semakin ketat. Maka dari itu, setiap perguruan tinggi dituntut untuk belomba-lomba memberikan kualitas terbaiknya guna meningkatkan *image* dan nama besarnya di kalangan masyarakat luas. Kesuksesan dan keunggulan sebuah perguruan tinggi sering dilihat dari keberhasilan prestasi belajar mahasiswa yang bernaung dalam perguruan tinggi tersebut. Sebaliknya, kegagalan dan rendahnya kualitas mahasiswa sering dilihat sebagai ketidakmampuan pihak perguruan tinggi menyelenggarakan proses kegiatan belajar mengajar dengan baik.

Untuk mendapatkan *output* mahasiswa berkualitas, perguruan tinggi tentunya juga harus dapat selektif menerima *input* calon mahasiswa yang memiliki peluang tinggi untuk sukses dalam menempuh pendidikannya. Proses seleksi penerimaan calon mahasiswa merupakan wadah yang tepat untuk menyaring calon mahasiswa yang layak diterima. Adapun salah satu langkah yang dibutuhkan dalam proses seleksi yakni dengan melakukan prediksi kinerja akademik calon mahasiswa. Prediksi kinerja akademik sangat penting untuk dilakukan karena dengan menganalisis pengalaman dan faktor-faktor lain dari calon mahasiswa, perguruan tinggi dapat melihat kemungkinan mahasiswa tersebut berpotensi memiliki keberhasilan dengan kinerja yang baik dalam menempuh studi atau tidak. Prediksi ini hanya sebagai salah satu bahan pertimbangan untuk penerimaan mahasiswa, karena disamping prediksi ini tentunya kemampuan mahasiswa dalam menerima ilmu akademik merupakan hal utama yang menjadi bahan pertimbangan.

Institut Teknologi Sepuluh Nopember (ITS) Surabaya sebagai salah satu perguruan tinggi ternama di Indonesia harus mempertahankan kualitas mahasiswanya. Salah satu seleksi penerimaan mahasiswa yang ada di ITS adalah Seleksi Penerimaan Calon Mahasiswa Program Studi Magister. Seleksi ini merupakan seleksi internal yang dilakukan ITS untuk menyaring calon mahasiswa baru yang ingin melanjutkan pendidikan akademiknya di jenjang Magister (S2). Seleksi internal ini dilakukan dengan melaksanakan tes kemampuan akademik bidang disiplin ilmu masing-masing dan tes wawancara yang dilengkapi dengan berkas-berkas penunjang seperti hasil TOEFL, TPA serta rencana penelitian.

Program Studi Magister Statistika sebagai salah satu prodi terbaik di ITS belum memiliki *tools* untuk melakukan prediksi kinerja mahasiswanya. *Tools* tersebut dibutuhkan sebagai salah satu bahan pertimbangan penerimaan calon mahasiswa. Maka dari itu, peneliti mengusulkan untuk membuat pemodelan prediksi kinerja mahasiswa yang nantinya dapat bermanfaat untuk pengelola Perguruan Tinggi, khususnya di Program Studi Magister Statistika ITS. Diharapkan dengan adanya sistem yang dapat memprediksi performansi mahasiswa tersebut mampu menunjang keputusan untuk seleksi penerimaan mahasiswa baru Program Studi Magister Statistika ITS ke depan.

Beberapa penyebab kegagalan mahasiswa yakni rendahnya kemampuan akademik, faktor pembiayaan, domisili saat menempuh studi dan faktor lainnya (Hastuti, K., 2012). *Database* perguruan tinggi menyimpan data akademik, administrasi dan biodata mahasiswa. Data tersebut apabila dieksplorasi dengan tepat maka dapat diketahui pola atau pengetahuan untuk mengambil sebuah keputusan (El-Halees, A., 2009). Dengan berbagai faktor baik data numerik, demografi, hingga data universitas asal calon mahasiswa yang digunakan sebagai acuan, diharapkan sistem dapat memberikan hasil berupa prediksi kinerja mahasiswa mengenai waktu kelulusan dan hasil nilai akademis yang didapatkan.

Dalam kasus prediksi, dibutuhkan sebuah metode atau model yang mampu menghasilkan pola klasifikasi data dengan tujuan akhir yaitu *forecasting*. Teknik atau metode yang akan digunakan untuk mengetahui informasi itu adalah dengan cara *mining* data akademik mahasiswa. *Data mining* terdiri berbagai teknik yang dapat digunakan untuk melakukan prediksi maupun klasifikasi, dimana teknik ini



memperkirakan kemungkinan yang akan terjadi di masa yang akan datang dengan melihat beberapa informasi dan pola data yang ada (Mustakim, 2015).

Dalam mengolah data mahasiswa yang berfokus terhadap prediksi, telah banyak dilakukan sebelumnya. Pertama, oleh David Hartanto Kamagi tentang implementasi data *mining* dengan Algoritma C4.5 (*Decision Tree*) untuk memprediksi tingkat kelulusan Mahasiswa. Hasil dari penelitian ini adalah sebuah sistem aplikasi berbasis *desktop* yang berhasil memprediksi kelulusan Mahasiswa dengan persentase 87,5% dari enam puluh data *training* dan empat puluh data *testing* (Kamagi, 2014). Kedua, penelitian oleh Mustafa, yaitu memprediksi kemungkinan mahasiswa baru untuk dapat menyelesaikan studi tepat waktu dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Dari hasil pengujian dengan menerapkan algoritma K-NN dan menggunakan data sampel alumni tahun wisuda 2004 sampai dengan 2010 untuk kasus lama dan data alumni tahun wisuda 2011 untuk kasus baru diperoleh tingkat akurasi sebesar 83,36%. (Mustafa, 2014).

Setiap metode yang digunakan pada penelitian-penelitian sebelumnya memiliki kelemahan dan kelebihan. Kelemahan mendasar dari metode yang dilakukan yaitu efisiensi waktu dalam pemrosesan data. Pengolahan data yang berukuran besar dengan banyak variabel independen yang digunakan mengakibatkan efisiensi metode kurang begitu baik. Adapun kelemahan yang lain adalah masih rendahnya tingkat akurasi yang diperoleh dari hasil pengujian. Sehingga guna pengembangan penelitian yang lebih baik, peneliti menggunakan metode *Support Vector Machine* (SVM) untuk penelitian prediksi kinerja mahasiswa yang dilakukan ini.

SVM dikenal sebagai salah satu metode klasifikasi yang memiliki tingkat akurasi yang tinggi dalam melakukan prediksi pengklasifikasian potensi pada data-data akademik. Terbukti dalam banyak implementasi, SVM memberikan hasil yang lebih baik dari algoritma lainnya, terutama dalam hal solusi yang dicapai. SVM menemukan solusi yang *global optimal*, sedangkan algoritma lain memiliki solusi *local optimal* (Santosa, 2007). Yang dimaksud *local optimal* adalah solusi yang dapat dicapai oleh sebuah algoritma namun berada dalam rentang nilai tertentu yang dibatasi, adapun *global optimal* berarti solusi yang dicapai berada untuk keseluruhan rentang input data. SVM juga memiliki kemampuan memecahkan masalah berdimensi tinggi walaupun dengan keterbatasan sampel (Zhang, 2010).

Pada kasus dalam penelitian ini, dibutuhkan penggunaan Metode *Multi-class SVM* untuk data demografi maupun data akademik mahasiswa yang bersifat non-linier. Penerapan algoritma SVM untuk data non-linier harus disertai dengan penggunaan fungsi Kernel untuk mendapatkan klasifikasi terbaik. Adapun akan dilakukan skenario pengujian dengan teknik *n-Folds Cross Validation* sehingga diharapkan sistem dapat memberikan hasil berupa prediksi kinerja dan keberhasilan akademik mahasiswa dengan tingkat akurasi yang tinggi.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang sudah diuraikan, permasalahan yang menjadi pokok bahasan kajian penelitian ini antara lain:

1. Apa saja faktor-faktor yang dapat mempengaruhi kinerja mahasiswa untuk dapat digunakan sebagai variabel independen pada penelitian.
2. Bagaimana cara penerapan algoritma *Support Vector Machine* untuk memprediksi kinerja mahasiswa.
3. Bagaimana analisa tingkat akurasi hasil prediksi kinerja mahasiswa menggunakan *Support Vector Machine* yang nantinya akan dibandingkan dengan data sebenarnya.

## **1.3. Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membuat model prediksi kinerja mahasiswa yang dapat digunakan sebagai salah satu pertimbangan evaluasi untuk pengelola program studi di perguruan tinggi dengan algoritma *Support Vector Machine*.
2. Mengevaluasi penerapan algoritma *Support Vector Machine* dalam pemodelan prediksi kinerja mahasiswa dengan mengukur akurasi hasil prediksi menggunakan teknik skenario uji *n-Folds Cross Validation*.

## **1.4. Batasan Penelitian**

Untuk lebih memfokuskan penelitian dan menyederhanakan permasalahan agar dapat diselesaikan dengan pendekatan metode ilmiah, peneliti menentukan ruang lingkup penelitian.

Batasan permasalahan yang digunakan dalam penelitian tesis ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian adalah seluruh data mahasiswa program studi Magister (S2) Statistika ITS Surabaya untuk Tahun Akademik 2011/2012 sampai dengan 2015/2016. Pertimbangan penggunaan keseluruhan data mahasiswa, baik yang sudah memiliki status akhir lulus maupun belum, dikarenakan agar sistem dapat menampilkan model prediksi yang akurat dengan tidak menghapus data sebenarnya.
2. Data mahasiswa yang belum memiliki status akhir, pada penelitian ini diberikan perlakuan khusus untuk memberi status akhir kelulusan dengan waktu tempuh lulus yang dirata-rata berdasarkan sisa dispensasi waktu studi mahasiswa tersebut.
3. Penentuan faktor-faktor yang mempengaruhi kinerja mahasiswa didapatkan berdasarkan hasil *focus group discussion* yang divalidasi oleh pakar/*decision maker* dalam hal ini adalah pengelola Program Studi. Faktor-faktor tersebut kemudian dilakukan uji korelasi untuk melihat variabel yang berpengaruh signifikan sehingga dapat dijadikan variabel independen untuk prediksi dengan SVM.

### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan dapat diperoleh dari penelitian ini meliputi manfaat praktis dan manfaat teoritis. Manfaat praktis adalah manfaat langsung dirasakan oleh praktisi pengguna hasil penelitian. Sedangkan manfaat teoritis berarti manfaat yang didapatkan oleh pembaca penelitian. Adapun manfaat tersebut dijabarkan sebagai berikut:

#### **a. Manfaat Praktis**

Penelitian ini diharapkan dapat memberikan manfaat sebagai salah satu pertimbangan pengambilan keputusan untuk seleksi penerimaan mahasiswa program studi Magister Statistika ITS Surabaya.

#### **b. Manfaat Teoritis**

Penelitian ini diharapkan dapat menjadi rujukan bagi penelitian-penelitian sejenis dan menambah keilmuan bagi penelitian-penelitian di bidang klasifikasi data *mining*.

*halaman ini sengaja dikosongkan*

## **BAB 2**

### **TINJAUAN PUSTAKA**

Pada bab ini akan dijelaskan mengenai kepustakaan yang mendukung penelitian antara lain mengenai Proses Klasifikasi, *Support Vector Machine*, Fungsi Kernel, Teknik Uji *n-Folds Cross Validation* dan studi literatur jurnal-jurnal sebelumnya yang dapat membantu penulis dalam menyelesaikan penelitian.

#### **2.1. Proses Klasifikasi**

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan kepada dokumen yang belum memiliki kategori (Goller, 2000). Mengklasifikasi dokumen merupakan salah satu cara untuk mengorganisasikan dokumen. Dokumen-dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan demikian, orang-orang yang melakukan pencarian informasi dapat dengan mudah melewati kategori yang tidak relevan dengan informasi yang dicari atau yang tidak menarik perhatian (Feldman, 2004). Pada penelitian ini, klasifikasi diterapkan untuk mengkategorikan data mahasiswa yang berpotensi memiliki kinerja akademik baik dan tidak baik.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Adapun dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu:

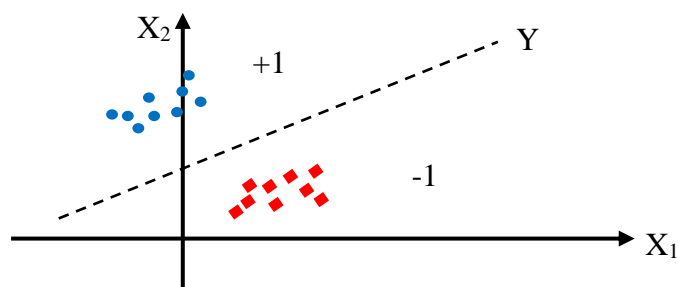
1. Proses *Training*

Pada proses *training* digunakan *training set* yang telah diketahui label-labelnya untuk membangun model atau fungsi.

2. Proses *Testing*

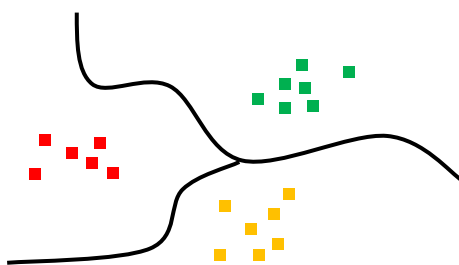
Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training*, maka digunakan data yang disebut dengan *testing set* untuk memprediksi label-labelnya.

Menurut Santosa (2007), klasifikasi dibagi menjadi dua macam yakni klasifikasi dua kelas dan multi kelas. Yang pertama, klasifikasi dua kelas dapat dijelaskan sebagai berikut. Misalkan kita memiliki set data *training*  $(x_i, y_i), i = 1, \dots, l$  dengan data input  $X = \{x_1, x_2, \dots, x_l\} \subseteq \mathbb{R}^N$  dan *output* yang bersangkutan  $Y = \{y_1, \dots, y_l\} \subseteq \{\pm 1\}^l$ . Tujuan dari klasifikasi dua kelas adalah menemukan suatu fungsi keputusan (*decision fuction*)  $f(x)$  yang secara akurat memprediksi kelas dari data *test*  $(x, y)$  yang berasal dari fungsi distribusi yang sama dengan data untuk *training*, lihat Gambar 2.1. Set data  $(x_i, y_i), i = 1, \dots, l$  biasa dinamakan *set training*, dimana  $x_i$  berkaitan dengan parameter *input* dan  $y_i$  menunjukkan parameter *output*.



**Gambar 2.1** Ilustrasi Klasifikasi Dua Kelas

Untuk klasifikasi Multi Kelas, misalkan kita memiliki *set* data untuk *training*  $(x_i, y_i), i = 1, \dots, l$  dengan data input  $X = \{x_1, x_2, \dots, x_l\} \subseteq \mathbb{R}^N$  dan *output* yang bersangkutan  $Y = \{y_1, \dots, y_l\} \subseteq \{1, 2, \dots, k\}^l$ . Terlihat bahwa *output*  $Y$  tidak lagi terbatas  $\pm 1$  seperti dalam kasus dua kelas. Output dari data kita bisa 1,2,3,4 atau bahkan 10. Ilustrasi ditampilkan dalam Gambar 2.2. Dalam hal ini kita harus mengelompokkan obyek yang kita pelajari ke dalam lebih dari dua kelas atau ke dalam  $k$  kelas dimana nilai  $k$  lebih dari 2. Dalam kasus demikian tugas klasifikasi menjadi lebih rumit dan perlu teknik khusus untuk mengatasinya.



**Gambar 2.2** Ilustrasi Klasifikasi Multi Kelas

Peningkatan jumlah dokumen yang begitu pesat telah mendorong berkembangnya metode pengklasifikasian secara otomatis. Metode-metode klasifikasi ini dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya. Keuntungan dari metode ini adalah dapat menghemat waktu kerja dan memperoleh hasil yang lebih baik.

## 2.2. *Support Vector Machine*

Pada kajian berikut akan dijelaskan mengenai dasar-dasar teori tentang metode yang akan digunakan dalam penelitian, yakni *Support Vector Machine* (SVM). Adapun penjelasan teoritis ini terdiri dari Definisi, Non Linier SVM, Karakteristik SVM, Multi kelas untuk SVM, Kelebihan dan Kekurangan penggunaan metode SVM.

### 2.2.1 Definisi *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar SVM sebenarnya merupakan kombinasi harmonis dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperplane*. Kernel diperkenalkan oleh Aronszajn tahun 1950, dan demikian juga dengan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkai komponen-komponen tersebut.

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. *Pattern* merupakan anggota dari dua buah kelas: +1 dan -1 dan berbagi alternatif garis pemisah (*discrimination boundaries*). *Margin* adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM (Christianini, 2000).

Data yang tersedia dinotasikan sebagai  $\vec{x}_i \in \mathcal{R}^d$  sedangkan label masing-masing dinotasikan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, \dots, l$  yang mana  $l$  adalah banyaknya data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi  $d$ , yang didefinisikan:

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.1)$$

Pattern  $\vec{x}_i$  yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan:

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2.2)$$

Sedangkan pattern  $\vec{x}_i$  yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan:

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (2.3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai *Quadratic Programming (QP) Problem*, yaitu mencari titik minimal persamaan (4) dengan memperhatikan *constraint* persamaan (8).

$$\min_w \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (2.4)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall_i \quad (2.5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya dengan *Lagrange Multiplier*.

$$L(w, b, a) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l a_i (y_i((x_i \cdot w_i + b) - 1))$$

$$(i = 1, 2, 3, \dots, l) \quad (2.6)$$

$a_i$  adalah *Lagrange multipliers*, yang bernilai nol atau positif ( $a_i \geq 0$ ). Nilai optimal dari Persamaan (6) dapat dihitung dengan meminimalkan L terhadap  $\vec{w}$  dan b, dan memaksimalkan L terhadap  $a_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient L = 0, persamaan (6) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung  $a_i$ , sebagaimana persamaan (7) berikut:

Maksimasi:

$$\sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j \vec{x}_i \vec{x}_j \quad (2.7)$$

Dengan *Constraint*:

$$a_i \geq 0 (i = 1, 2, 3, \dots, l) \quad \sum_{i=1}^l a_i y_i = 0 \quad (2.8)$$



Dari hasil perhitungan ini diperoleh  $a_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $a_i$  yang positif inilah yang disebut sebagai *support vector* (Vapnik, 1995).

Penjelasan di atas berdasarkan asumsi bahwa kedua belah kelas dapat terpisah secara sempurna oleh *hyperplane* (*linear separable*). Akan tetapi, pada umumnya dua belah kelas pada *input space* tidak dapat terpisah secara sempurna (*non linear separable*). Hal ini menyebabkan constraint pada persamaan (8) tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan. Untuk mengatasi masalah ini, SVM dirumuskan ulang dengan memperkenalkan teknik *softmargin*.

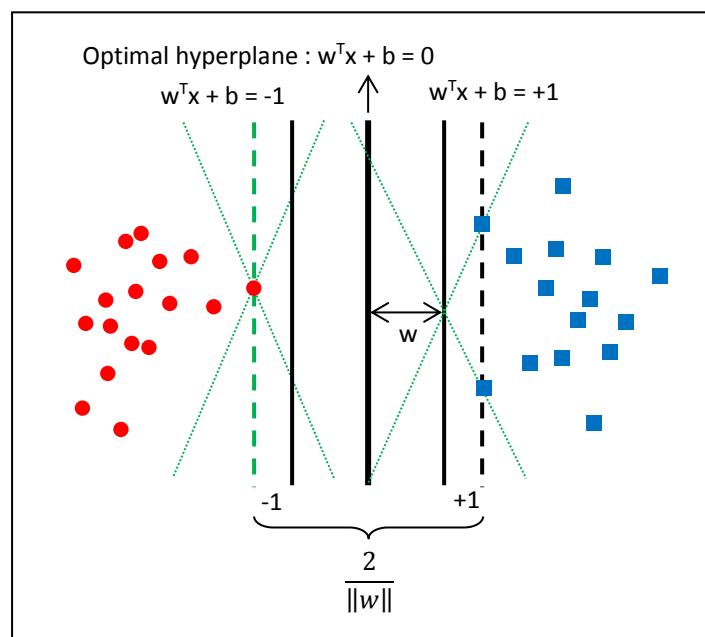
Dalam *softmargin*, Persamaan (5) dimodifikasi dengan memasukkan *slack variable*  $\xi$  ( $\xi > 0$ ) sebagai berikut:

$$y_i(x_i \cdot w + b) \geq 1 - \xi_i \forall_i \quad (2.9)$$

Dengan demikian persamaan (4) diubah menjadi:

$$\min_{\bar{w}} \tau(w, \xi) = \frac{1}{2} \|\bar{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (2.10)$$

Parameter C dipilih untuk mengontrol *tradeoff* antara margin dan error klasifikasi  $\xi$ . Nilai C yang besar berarti akan memberikan *penalty* yang lebih besar terhadap error klasifikasi tersebut (Sembiring, 2007).

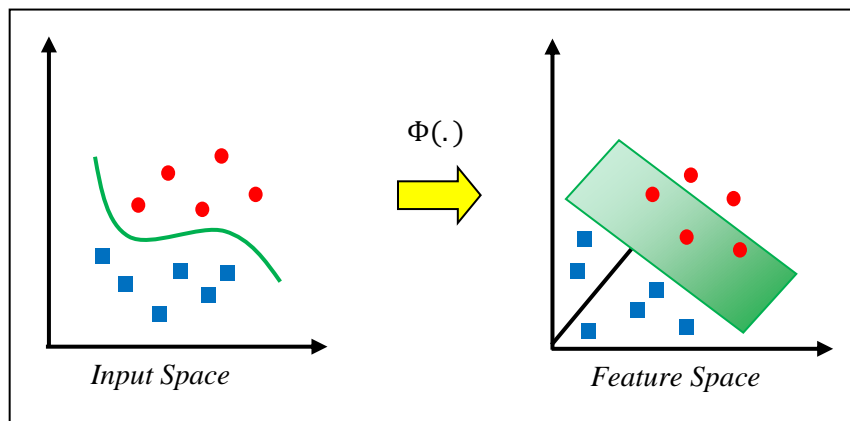


**Gambar 2.3** Ilustrasi Mencari Fungsi Pemisah Optimal dalam Kasus Klasifikasi

### 2.2.2. Non-Linear Classification (Klasifikasi yang tidak Linier)

Pada umumnya masalah dalam domain dunia nyata (*real world problem*) jarang yang bersifat linear *separable* (tidak terpisahkan secara linear), tetapi bersifat non-linear. Untuk menyelesaikan problem non-linear, SVM dimodifikasi dengan memasukkan fungsi kernel. Dalam non-linear SVM, pertama-tama data  $\vec{x}$  dipetakan oleh fungsi  $\Phi(\vec{x})$  ke ruang vektor yang berdimensi lebih tinggi. *Hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan. Selanjutnya gambar (2.1) menunjukkan bahwa fungsi  $\Phi$  memetakan tiap data pada input space tersebut ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), sehingga kedua kelas dapat dipisahkan secara linear oleh sebuah *hyperplane*. Notasi matematika dari *mapping* ini adalah sebagai berikut:

$$\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^d, d < q \quad (2.11)$$



**Gambar 2.4** Fungsi  $\Phi$  memetakan data ke ruang *vector* yang berdimensi lebih tinggi

Selanjutnya proses pembelajaran pada SVM dalam menemukan titik-titik *support vector*, hanya bergantung pada dot product dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu  $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$ .

Karena umumnya transformasi  $\Phi$  ini tidak diketahui, dan sangat sulit untuk dipahami secara mudah, maka perhitungan *dot product* dapat digantikan dengan fungsi Kernel  $K(\Phi(\vec{x}_i), \Phi(\vec{x}_j))$  yang mendefinisikan secara implisit transformasi  $\Phi$ . Hal ini disebut sebagai *Kernel Trick*, yang dirumuskan:

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \quad (2.12)$$

$$\begin{aligned}
f(\Phi(\vec{x})) &= \vec{w} \cdot \Phi(\vec{x}) + b \\
&= \sum_{i=1, X, eSV}^n a_i y_i \Phi(\vec{x}) \cdot \Phi(\vec{x}_i) + b \\
&= \sum_{i=1, X, eSV}^n a_i y_i K(x, x_i) + b
\end{aligned} \tag{2.13}$$

SV pada persamaan di atas dimaksudkan dengan subset dari *training set* yang terpilih sebagai *support vector*, dengan kata lain  $\vec{x}_i$  data yang berkorespondensi pada  $a_i \geq 0$ .

### 2.2.3. Karakteristik Support Vector Machine (SVM)

Menurut (Nugroho, dkk, 2003), karakteristik SVM secara umum dirangkumkan sebagai berikut:

1. Secara prinsip SVM adalah *linear classifier*.
2. *Pattern recognition* dilakukan dengan mentransformasikan data pada *input space* ke ruang yang berdimensi lebih tinggi, dan optimisasi dilakukan pada ruang *vector* yang baru tersebut. Hal ini membedakan SVM dari solusi *pattern recognition* pada umumnya, yang melakukan optimisasi parameter pada ruang hasil transformasi yang berdimensi lebih rendah daripada dimensi *input space*.
3. Menerapkan strategi *Structural Risk Minimization (SRM)*.
4. Prinsip kerja SVM pada dasarnya hanya mampu menangani klasifikasi dua kelas.

### 2.2.4 Multiclass Support Vector Machine (SVM)

Ada dua pilihan untuk mengimplementasikan *multiclass SVM* yaitu dengan menggabungkan beberapa SVM biner atau menggabungkan semua data yang terdiri dari beberapa kelas ke dalam sebuah bentuk permasalahan optimal. Namun pada pendekatan yang kedua permasalahan optimasi yang harus diselesaikan jauh lebih rumit. Berikut ini adalah metode yang umum digunakan untuk mengimplementasikan *multiclass SVM* dengan pendekatan yang pertama:

1. *Metode one-against-all* (satu lawan semua)

Dengan menggunakan metode ini, dibangun k buah model SVM biner (k adalah jumlah kelas)

## 2. Metode *one-against-one* (satu lawan satu)

Dengan menggunakan metode ini, dibangun  $k(k-1)/2$  buah model klasifikasi biner ( $k$  adalah jumlah kelas). Terdapat beberapa metode untuk melakukan pengujian setelah keseluruhan  $k(k-1)/2$  model klasifikasi selesai dibangun. Salah satunya adalah metode *voting* (Santosa, 2007).

### 2.2.5 Kelebihan *Support Vector Machine* (SVM)

Adapun beberapa keuntungan dari metode SVM adalah sebagai berikut:

#### 1. Generalisasi

Generalisasi didefinisikan sebagai kemampuan suatu metode untuk mengklasifikasikan suatu pattern, yang tidak termasuk data yang dipakai dalam fase pembelajaran metode itu.

#### 2. *Curse of dimensionality*

*Curse of dimensionality* didefinisikan sebagai masalah yang dihadapi suatu metode pattern recognition dalam mengestimasi parameter dikarenakan jumlah sampel data yang relatif lebih sedikit dibandingkan dengan dimensional ruang vektor data tersebut.

#### 3. *Feasibility*

SVM dapat diimplementasikan relatif mudah, karena proses penentuan *support vector* dapat dirumuskan dalam QP problem (Nugroho, dkk, 2003)

### 2.2.6 Kekurangan *Support Vector Machine* (SVM)

Adapun beberapa kerugian dari metode SVM adalah sebagai berikut:

1. Sulit dipakai problem berskala besar. Dalam hal ini dimaksudkan dengan jumlah sampel yang diolah.

2. SVM secara teoritik dikembangkan untuk problem klasifikasi dengan dua kelas. Dewasa ini SVM telah dimodifikasi agar dapat menyelesaikan masalah dengan lebih dari dua kelas (Nugroho, dkk, 2003).

## 2.3. Fungsi Kernel

Pada mulanya klasifikasi dikembangkan dengan asumsi kelinieran. Sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linear saja. Akan tetapi untuk

menghadapi kasus yang tidak linier maka dapat menggunakan bantuan berbagai macam fungsi Kernel. Kernel *trick* memberikan berbagai kemudahan, karena dalam proses pembelajaran SVM. Untuk menentukan *support vector*, maka cukup dengan mengetahui fungsi Kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi non-linear.

Menurut (Karatzoglou, dkk, 2004) ada beberapa fungsi Kernel yang sering digunakan dalam literatur SVM antara lain sebagai berikut:

- a. Kernel *linear* adalah kernel yang paling sederhana dari semua fungsi kernel. Kernel ini biasa digunakan dalam kasus klasifikasi teks.

$$K(x_i, x_j) = x_i^T x_j \quad (2.14)$$

- b. Kernel *Polynomial* adalah kernel yang sering digunakan untuk klasifikasi gambar.

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^p, \gamma > 0 \quad (2.15)$$

- c. Kernel *Radial Basis Gaussian* adalah kernel yang umum digunakan untuk data yang sudah valid (*available*) dan merupakan default dalam tools SVM.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (2.16)$$

- d. Kernel *Tangent Hyperbolic* adalah kernel yang sering digunakan untuk *neural networks*.

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (2.17)$$

Pemilihan fungsi Kernel yang tepat merupakan hal yang sangat penting karena akan menentukan *feature space*, dimana fungsi *classifier* akan dicari. Sepanjang fungsi kernelnya sesuai/cocok, SVM akan beroperasi secara benar meskipun tidak tahu pemetaan yang digunakan (Santosa, 2007; Robandi & Prasetyo, 2008). Menurut Scholkopf dan Smola (1997), fungsi Kernel Gaussian RBF memiliki kelebihan yaitu secara otomatis menentukan nilai rentang tak terhingga. Gaussian RBF juga efektif menghindari *overfitting* dengan memilih nilai yang tepat untuk parameter C dan  $\gamma$  dan RBF yang baik digunakan ketika tidak ada pengetahuan terdahulu. Fungsi kernel yang direkomendasikan adalah fungsi kernel RBF karena dapat memetakan hubungan tidak linier, RBF lebih *robust* terhadap *outlier* karena fungsi kernel RBF berada pada selang  $(-\infty, \infty)$  sedangkan fungsi Kernel yang lain memiliki rentang antara (-1 sampai dengan 1) (Hsu, Chang, & Lin, 2003).

#### 2.4. Regresi Logistik sebagai Metode Pembandingan

Analisis Regresi dalam statistika adalah salah satu metode untuk menentukan hubungan sebab-akibat antara satu variabel dengan variabel-variabel yang lain. Variabel “penyebab” disebut dengan bermacam-macam istilah, di antaranya seperti variabel penjelas, variabel independen, atau secara bebas, variabel  $x$  (karena seringkali digambarkan dalam grafik sebagai absis, atau sumbu  $x$ ). Variabel “terkena akibat” dikenal sebagai variabel yang dipengaruhi, variabel dependen, variabel terikat, atau variabel  $Y$ . Kedua variabel ini dapat merupakan variabel acak (*random*), namun variabel yang dipengaruhi harus selalu variabel acak. Analisis Regresi adalah salah satu analisis yang paling populer dan luas pemakaiannya. Hampir semua bidang ilmu yang memerlukan analisis sebab-akibat boleh dipastikan mengenal analisis ini.

Adapun Regresi Logistik (kadang disebut model logistik atau model logit) merupakan salah satu bagian dari analisis regresi yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa, dengan mencocokkan data pada fungsi logit kurva logistik. Metode ini merupakan model linier umum yang digunakan untuk regresi binomial. Seperti analisis regresi pada umumnya, metode ini menggunakan beberapa variabel bebas, baik numerik maupun kategorik. Misalnya, probabilitas bahwa orang yang menderita serangan jantung pada waktu tertentu dapat diprediksi dari informasi usia, jenis kelamin, dan indeks masa tubuh. Regresi Logistik juga digunakan secara luas pada bidang kedokteran, ilmu sosial, dan bahkan pada bidang pemasaran, seperti prediksi kecenderungan pelanggan untuk membeli suatu produk atau berhenti berlangganan. Regresi Logistik tidak memerlukan asumsi normalitas, heteroskedastisitas, dan autokorelasi, dikarenakan variabel terikat yang terdapat pada Regresi Logistik merupakan variabel *dummy* (0 dan 1), sehingga residualnya, tidak memerlukan ketiga pengujian tersebut. Untuk asumsi multikolinearitas, karena hanya melibatkan variabel-variabel bebas, maka masih perlu untuk dilakukan pengujian. Untuk pengujian multikolinieritas ini dapat digunakan uji kebaikan suai (*goodness of fit test*), yang kemudian dilanjutkan dengan pengujian hipotesis guna melihat variabel-variabel bebas mana saja yang signifikan, sehingga dapat tetap digunakan dalam penelitian. Selanjutnya, diantara variabel-variabel bebas yang signifikan, dapat dibentuk suatu matriks korelasi, dan apabila tidak terdapat variabel-variabel bebas yang saling memiliki

korelasi yang tinggi, maka dapat disimpulkan bahwa tidak terdapat gangguan multikolinieritas pada model penelitian (Hosmer, 2011)

Dalam penelitian ini, metode Regresi Logistik digunakan sebagai metode pembandingan yang akan diukur akurasi klasifikasinya pada prediksi performansi mahasiswa. Hal tersebut dilakukan guna mengetahui perbandingan penggunaan metode *Support Vector Machine* yang diajukan dalam penelitian dan Regresi Logistik sebagai metode statistika konvensional.

## 2.5. Evaluasi Performansi Metode Klasifikasi

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Nugroho, dkk, 2003). Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Permasalahan pada klasifikasi biner, akurasi klasifikasi disajikan pada Tabel 2.1.

**Tabel 2.1** Akurasi Klasifikasi

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan :

TP : *True Positive* (Jumlah prediksi benar pada kelas positif)

FP : *False Positive* (Jumlah prediksi salah pada kelas positif)

FN : *False Negative* (Jumlah prediksi salah pada kelas negatif)

TN : *True Negative* (Jumlah prediksi benar pada kelas negatif)

Pada Tabel 2.1 nilai TP (*true positive*) dan TN (*true negative*) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, presisi, dan recall. Jika label prediksi keluaran bernilai benar (*true*) dan nilai sebenarnya bernilai salah (*false*) disebut sebagai *false positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*false*) dan nilai sebenarnya bernilai benar (*true*) maka hal ini disebut sebagai *false negative* (FN). Berdasarkan Tabel 2.1 perhitungan evaluasi performansi metode klasifikasi dapat dilakukan dengan formula sebagai berikut:

$$\text{Akurasi} = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (2.18)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (2.19)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (2.20)$$

## 2.6. *N-Folds Cross Validation*

*N-folds Cross Validation* adalah salah satu teknik untuk validasi yang sangat populer digunakan. Metode Validasi dengan *n-folds* sangat cocok digunakan untuk kasus data yang jumlah sampelnya terbatas. Untuk melakukan proses klasifikasi tentunya data dibagi ke dalam *training* dan *testing*. Ketika data yang digunakan untuk *training* sangat sedikit kemungkinan adalah data yang digunakan sangat representatif. Dalam *n-folds cross validation*, data (D) dibagi ke dalam *n subsets*  $D_1, D_2, D_3, \dots, D_n$  dengan jumlah yang sama. Data yang digunakan untuk *training* adalah *subsets* data *n-1* yang dikombinasikan secara bersama-sama dan kemudian diaplikasikan untuk sisa satu *subsets* data sebagai hasil *testing*. Proses ini diulangi sebanyak *n subsets* dan hasil akurasi klasifikasi yaitu hasil rata-rata dari setiap data *training* dan *testing*. *N-folds* yang biasa digunakan adalah 3, 5, 10 dan 20 (Bolon, Sanchez & Alonso, 2015).

## 2.7. Penelitian Terdahulu

Untuk memperjelas posisi penelitian ini, maka disusun suatu tinjauan ulang terhadap penelitian-penelitian sejenis yang telah dilakukan sebelumnya dengan berdasarkan pada pendekatan tinjauan dari kesamaan obyek bahasan maupun dengan kesamaan metode yang digunakan.

### 2.7.1. Penelitian yang Terkait dengan Prediksi Kinerja Akademik

Pada bagian ini akan dibahas mengenai kepustakaan penelitian terdahulu terkait dengan prediksi kinerja akademik. Penulis mencoba untuk mencari literatur tentang faktor-faktor yang dapat dijadikan variabel independen untuk klasifikasi prediksi kinerja dan keberhasilan mahasiswa yang dibahas dalam penelitian ini. Berikut adalah beberapa jurnal terkait dengan prediksi kinerja akademik.

Penelitian yang dilakukan oleh Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan (2015) berjudul *Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector*. Penelitian yang dilakukan peneliti dari



Punjab *Technical University* India tersebut membahas tentang bagaimana mengidentifikasi performansi siswa setingkat SMA dari segi kelambanan belajar di kalangan pelajar dan menampilkan prediksi menggunakan teknik klasifikasi data *mining* dengan berbagai macam algoritma. Data *real* dari sejumlah sekolah setingkat SMA di India diambil untuk dilakukan *filter* sesuai dengan variabel potensial yang diinginkan lalu kemudian diolah menggunakan *open source software* WEKA. Dataset catatan akademik para pelajar yang sudah dikumpulkan diuji dan diterapkan pada berbagai algoritma klasifikasi seperti *Multi Layer Perception*, *Naive Bayes Classifier*, *SMO*, *J48* dan *REPTree* dengan perangkat lunak WEKA.

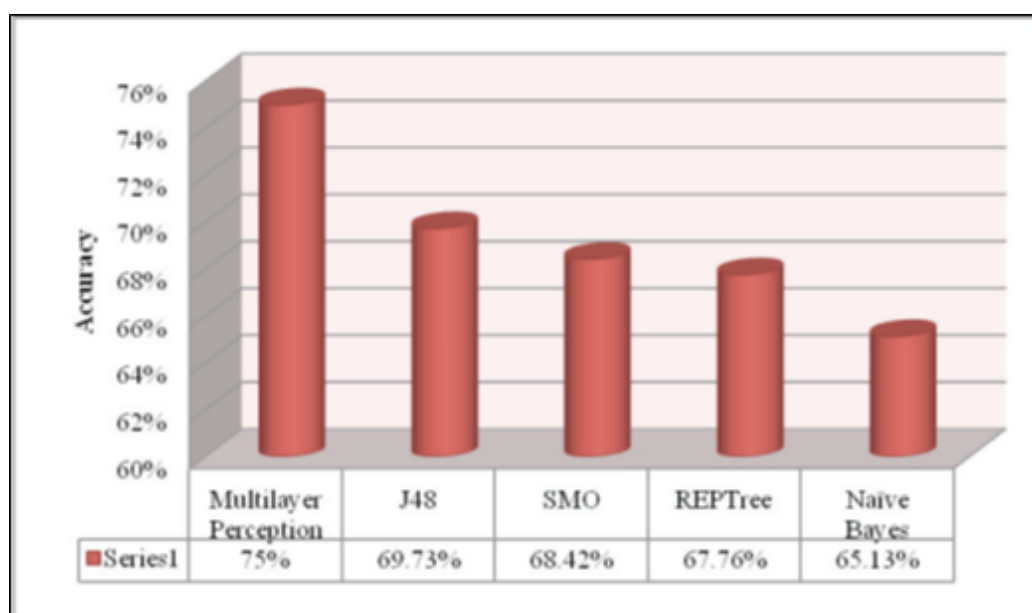
Sebanyak 152 data digunakan untuk analisis penelitian ini. Dalam penelitian ini, pemilihan variabel potensial yang berpengaruh untuk klasifikasi dipilih menggunakan analisis variabel yang ada dalam fasilitas WEKA antara lain metode *Chi Squared* atribut, *Info Gain* atribut, *Symmetrical Uncert* atribut dan *ReliefF* atribut. Setelah itu dilakukan evaluasi ranking/peringkat variabel sesuai dengan yang paling berpengaruh. Perangkat lunak WEKA memiliki menu Metode *Ranker Search Technique* untuk evaluasi ranking variabel tersebut. Berikut Tabel 2.2 merupakan hasil pemilihan variabel potensial yang sudah dievaluasi dengan metode *Ranker Search Technique*.

**Tabel 2.2** Variabel dengan Potensial Tinggi (Kaur, P. dkk., 2015)

<b>Variable Name</b>	<b>Description</b>	<b>Rank Values</b>
INT-GR	Internal grade of student	1.65
ATDN	Attendance count	2.2225
SEX	Student's Sex	3.6
PTUI	Private tuition	3.525
MOB	Student having mobile	5.375
INS-HIGH	Institution at high level	5.925
COM-HM	Computer at home	8.325
NET-ACS	Student having net access	9.2

Setelah melakukan analisis variabel yang berpotensi mempengaruhi proses klasifikasi, selanjutnya dilakukan teknik klasifikasi dengan berbagai macam algoritma yang sudah disebutkan sebelumnya. Hasil akhir yang didapatkan pada penelitian tersebut

berupa tingkat akurasi prediksi berdasarkan semua algoritma klasifikasi dan perbandingannya untuk menemukan kinerja algoritma klasifikasi terbaik di antara semua. Penelitian ini menampilkan pentingnya Prediksi dan Klasifikasi berdasarkan algoritma data mining di bidang pendidikan dan juga menyajikan hasil terbaiknya dari beberapa klasifikasi yang dilakukan. Adapun hasil pengujian akurasi masing-masing algoritma ditampilkan dalam Gambar 2.5.



**Gambar 2.5** Komparasi Uji Akurasi Masing-masing Algoritma (Kaur, P. dkk., 2015)

Di antara semua data hasil uji akurasi, metode *Multi Layer Perception* (MLP) memiliki tingkat akurasi tertinggi yakni 75%. Dengan demikian MLP terbukti menjadi algoritma *classifier* yang paling efektif dan efisien. Penelitian membantu pihak sekolah untuk mengidentifikasi siswa yang memiliki kelambanan dalam proses belajar sehingga hasil ini dapat memberikan dasar untuk memutuskan bantuan khusus kepada siswa tersebut.

Penelitian yang dilakukan oleh Mashael A. Al-Barrak dan Muna Al-Razgan (2016) yang berjudul *Predicting Students Final GPA Using Decision Trees: A Case Study*. Penelitian ini merupakan salah satu penelitian yang masuk dalam *International Journal of Information and Education Technology*, Vol. 6, No. 7, July 2016. Dalam penelitian ini, penulis menggunakan data pendidikan untuk memprediksi IPK akhir mahasiswa berdasarkan nilai sebelumnya pada setiap mata kuliah. Algoritma yang diterapkan untuk

prediksi IPK ini adalah algoritma *Decision Tree* J48. Penelitian ini juga mengevaluasi hasil yang didapat dari IPK untuk mengidentifikasi mata kuliah paling penting dan berpengaruh untuk rencana studi ke depan berdasarkan nilai-nilai mata kuliah wajib dari mahasiswa tersebut.

Dalam penelitian ini, peneliti menganalisis Data yang diperoleh dari Departemen Teknologi Informasi dengan bidang Ilmu Komputer di Universitas King Saud, Riyadh. Data transkrip di tahun 2012 dikumpulkan dalam sistem *database* dengan jumlah total 236 mahasiswa. Data tersebut dikumpulkan dalam Microsoft Excel. Setiap mahasiswa memiliki atribut antara lain nama mahasiswa, ID mahasiswa, IPK, sks lulus, kebangsaan, kampus, dan semua mata kuliah yang diambil oleh mahasiswa. Sama seperti penelitian sebelumnya, perangkat lunak yang digunakan untuk olah data adalah *software* WEKA.

Sebagai hasil kebermanfaatan dari penelitian ini, akan menghasilkan informasi untuk mahasiswa dan manajemen perguruan tinggi dalam pengambilan tindakan yang tepat untuk meningkatkan hasil IPK mahasiswa, dan berkontribusi untuk pendidikan yang lebih berkualitas. Adapun untuk penelitian selanjutnya, penulis mengusulkan untuk melakukan generalisasi penelitian dan menambahkan mata kuliah pilihan dan umum untuk mendapatkan hasil yang lebih akurat. Penulis juga mengusulkan untuk prediksi dan klasifikasi IPK dilakukan dengan teknik data mining lainnya, seperti *Neural Network* dan *clustering*.

Penelitian yang dilakukan oleh O. C. Asogwa dan A. V. Oladugba (2015) dengan judul *Of Students Academic Performance Rates Using Artificial Neural Networks* (ANNs). Penelitian ini dipublikasi pada *American Journal of Applied Mathematics and Statistics* 2015. Penelitian ini menunjukkan bahwa model berbasis Jaringan Syaraf Tiruan atau dikenal dengan *Artificial Neural Networks* (ANN) mampu memprediksi kinerja mahasiswa di semester satu dengan akurasi yang tinggi. Penggunaan *Multiple Feed-forward Neural Networks* untuk memprediksi prestasi akhir mahasiswa dengan pembagian klasifikasi dua kelas.

Data yang digunakan dalam penelitian ini adalah data sekunder yang dikumpulkan dari file mahasiswa fakultas kedokteran hewan, *University of Ibadan* pada tahun 2009 hingga 2013. *Record* semua mahasiswa dengan bidang kosong atau informasi yang tidak lengkap telah dihapus dari data yang dikumpulkan. Sejumlah total 420 data mahasiswa dikumpulkan dan digunakan untuk penelitian ini. Melalui tinjauan

dari literatur, sejumlah faktor terkait sosial-ekonomi, biologis, lingkungan, akademik, dan lainnya seperti persyaratan masuk dianggap memiliki pengaruh pada tingkat kinerja performansi akademik mahasiswa. Berikut Tabel 2.3 merupakan daftar variabel yang digunakan guna mengimplementasikan algoritma ANNs dalam memprediksi kinerja mahasiswa pada penelitian ini.

**Tabel 2.3** Variabel yang digunakan untuk Prediksi  
(Asogwa, O. C. dan Oladugba, A. V., 2015)

No.	Variables descriptions	Data type	Location	Code
1	Gender	Categorical	Input	1=male, 0=female
2	Parents Marital Status	Categorical	Input	0= Married and living together; 1= Father die; 2= Mother die; 3= Mother & father die; 4= Married but living apart because of occupation; 5= Married but living apart because other reasons; 6 = Divorced
3	Father's Education	Categorical	Input	0 = Uneducated; 1= Primary school; 2 = Secondary school; 3 = Degree; 4 = Vocational; 5 = Diploma; 6 = Masters degree; 7= Doctors degree
4	Mother's Education	Categorical	Input	0 = Uneducated 1 = Primary school 2 = Secondary school 3 = Degree 4 = Vocational 5 = Diploma 6 = Masters degree 7 = Doctors degree
5	Fathers Occupation	Categorical	Input	0 = Government service (govt. officer) 1 = Government service (temporary) 2 = State enterprise employee 3 = Company employee 4 = Ownership 5 = Farmer, fisher 6 = Freelance 7 = Unemployed 8 = Others
6	Mother's Occupation	Categorical	Input	0 = Government service (govt. officer) 1 = Government service (temporary) 2 = State enterprise employee 3 = Company employee 4 = Ownership 5 = Farmer, fisher 6 = Freelance 7 = Unemployed 8 = Others
8	Age at Entry	Categorical	Input	0 = Below 23 years 1 = 23 years and above
9	Time delay before admission	Categorical	Input	0=1year 1=2years 2=3years and above
10	Type Of Secondary School Attended	Categorical	Input	0= Private 1= State 2= Federal
11	Location of Secondary School attended	Categorical	Input	0=Urban 1= Rural

12	Post UTME	Numerical	Input	True value
13	Physically impaired	Categorical	Input	0 = Disable 2 = Not disable
14	Performance outcome	Categorical	output	1 = Promoted students 2 = Repeating students 3 = Demoted students

Faktor-faktor pada Tabel 2.3 tersebut kemudian dilakukan pemodelan *Artificial Neural Network*. Faktor-faktor yang mempengaruhi dikategorikan sebagai variabel *input* (unit). Variabel *output* pada sisi lain akan mewakili tingkat kinerja performansi mahasiswa. *Output* tersebut diklasifikasikan menjadi tiga kategori, yakni mahasiswa yang wisuda tepat waktu, mahasiswa yang harus mengulang semester, dan mahasiswa yang dikeluarkan (DO). Adapun *Software* Matlab R2009a digunakan sebagai alat statistik untuk analisis. Output dari klasifikasi menggunakan metode ANNs tersebut kemudian dibandingkan dengan data klasifikasi yang sebenarnya dan dievaluasi akurasi. Perbandingan diringkas dalam matriks yang ditampilkan dalam Tabel 2.4 berikut.

**Tabel 2.4** Perbandingan Hasil Klasifikasi dengan data sebenarnya  
(Asogwa, O. C. dan Oladugba, A. V., 2015)

Students Performances	<b>Promoted</b>	<b>Repeated</b>	<b>Demoted</b>
<b>Promoted</b>	83 (100%)	0 (0%)	0 (0%)
<b>Repeated</b>	6 (37.5%)	10 (62.5%)	0 (0%)
<b>Demoted</b>	2 (6.5%)	0 (0%)	29 (93.5%)
<b>Mean Correct Classification Rate CCR = 97.07%</b>			

Dari Tabel 2.4, setiap baris mewakili klasifikasi dari data sebenarnya dan setiap kolom merupakan hasil data yang dievaluasi oleh *classifier*. Dengan penggunaan metode ANN ini mampu memprediksi secara akurat 83 dari 83 untuk mahasiswa yang wisuda tepat waktu pada tahun keempat, 10 dari 16 untuk mahasiswa dinyatakan harus mengulang, dan 29 dari 31 untuk mahasiswa DO. Hal ini memberikan akurasi 100% untuk klasifikasi kelas pertama, 62,5% untuk kelas kedua, dan 93,5% klasifikasi kategori DO. Hal ini menunjukkan kinerja yang baik dengan akurasi klasifikasi mencapai sekitar 97,07% untuk metode *Artificial Neural Network* Model pada penelitian ini.

Penelitian yang dilakukan oleh Saeed Hosseini Teshnizi, dkk (2015) dengan judul *A Comparison of Logistic Regression Model and Artificial Neural Networks in Predicting of Student's Academic Failure*. Penelitian ini mengkaji tentang komparasi atau perbandingan klasifikasi antara Regresi Logistik dan *Artificial Neural Networks* (ANNs) yang diterapkan pada kasus prediksi kegagalan mahasiswa di bidang akademiknya. Data yang digunakan untuk penelitian ini meliputi 18 pertanyaan tentang situasi studi dari 275 mahasiswa yang dipilih secara acak dari beberapa jurusan yakni keperawatan & kebidanan serta jurusan ilmu kedokteran di Universitas Hormozgan pada tahun 2013. Regresi Logistik dengan metode *forward* dan *Artificial Neural Network* dengan 15 *neuron hidden layer* diterapkan di data *set*. Akurasi dari model dalam memprediksi kegagalan akademis dibandingkan dengan menggunakan metode *Receiver Operating Characteristic* (ROC) dan akurasi klasifikasi. Dalam penelitian ini, Indeks Prestasi Kumulatif (IPK) dari semester sebelumnya adalah *output* atau variabel respon yakni variabel yang mewakili kinerja mahasiswa, dan variabel lain seperti yang ditunjukkan di Tabel 2.5 berikut digunakan sebagai *input* atau variabel independen.

**Tabel 2.5** Karakteristik Mahasiswa untuk Prediksi Kegagalan Akademis (Teshnizi, S. H., dkk., 2015)

<b>Variable</b>	<b>Values</b>
Age (Year)	Less than 20, 20 to 25, More than 25
Gender	Male, Female
Place of residence	Home, Dormitory
Marital status	Married, Single
School	Nursing & Midwifery, Paramedic
Term	1, 2, ..., $\geq 10$
Grade Point Average (GPA)	Score between 0 to 20
Father's education	Under high school diploma, Diploma, Upper diploma
Mother's education	Under high school diploma, Diploma, Upper diploma
Guest student	Yes, No
Transitional Student	Yes, No
At least once failed of a course	Yes, No
Extracurricular activities	Yes, No

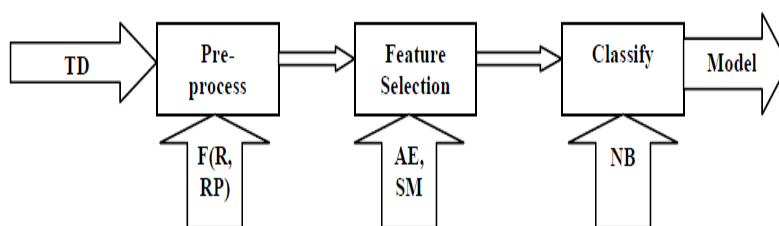
Employment status	Yes, No
Interest in the field	Very much, Much, Moderate, Little, Very little
GPA of the previous semester	Score between 0 to 20

Secara signifikan, AUROC untuk ANN (15) Model (89%) adalah jauh lebih tinggi dari pada model Regresi Logistik (55%). Adapun dengan sensitivitas untuk dua model menunjukkan bahwa akurasi klasifikasi kegagalan akademik menggunakan ANN (76,6%) lebih akurat dibandingkan dengan Model Regresi Logistik (67,3%). Secara umum, hasil penelitian ini menunjukkan bahwa di antara 9 *neuron* ANNs, dengan 15 neuron di *hidden layer* memiliki kinerja yang lebih baik. Dibandingkan dengan konvensional Model Regresi Logistik, model ANN dalam penelitian ini lebih akurat dalam memprediksi kegagalan akademis dan memiliki kinerja secara keseluruhan lebih baik dengan indeks yang lebih tinggi. Oleh karena itu, berdasarkan hasil lainnya dan data kegagalan akademis, tampaknya untuk variabel dependen dalam kasus ini, metode *Artificial Neural Network* adalah metode yang tepat untuk digunakan.

Penelitian yang dilakukan oleh Humera Shaziya, dkk (2015) dengan judul *Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier*. Penelitian ini mengkaji tentang prediksi kinerja mahasiswa dalam ujian akhir semester. Pendekatan ini didasarkan pada algoritma Klasifikasi *Naive Bayes*. Tujuannya adalah untuk memprediksi hasil nilai mahasiswa pada akhir semester mereka. Ini membantu institusi pendidikan baik dosen dan mahasiswa serta semua pihak yang terlibat dalam sistem akademik. Mahasiswa dan dosen dapat mengambil tindakan yang diperlukan untuk meningkatkan hasil nilai akhir semester dari prediksi yang dilihat. Sebuah data *set* pelatihan mahasiswa diambil untuk membangun model *Naive Bayes*. Model ini kemudian diterapkan pada data uji untuk memprediksi hasil ujian akhir semester para mahasiswa. Dalam penelitian ini, tingkatan dari mahasiswa berpengaruh terhadap jumlah atribut yang akan digunakan sebagai bahan dasar penelitian. Data yang dipilih untuk prediksi mahasiswa terdiri dari delapan atribut. Tabel 2.6 menunjukkan atribut yang berbeda dengan deskripsi dan nilai-nilainya.

**Tabel 2.6** Deskripsi Atribut dari Data *Set* (Shaziya, H., dkk., 2015)

No	Attribute	Description	Description Values
1.	PSM	Previous Semester Marks	First > 60%; Second > 45% & < 65%; Third > 35% & < 45%
2.	CTG	Class Test Grade	Poor; Average; Good
3.	SEM	Seminar Performance	Poor; Average; Good
4.	ASS	Assignment	Yes; No
5.	GP	General Proficiency	Yes; No
6.	ATT	Attendance	Poor; Average; Good
7.	LW	Lab Work	Yes; No
8.	ESM	End Semester Marks	First > 60%; Second >45% & < 65%; Third > 35% & < 45%



**Gambar 2.6** Model Proses Prediksi Ujian Akhir Semester Mahasiswa (Shaziya, H., dkk., 2015)

Adapun atribut-atribut tersebut kemudian diolah dengan Weka 3.7.10 menggunakan metode *Naive Bayes*. Proses *Pre-Processing* Data diperlukan untuk *filter* atribut data yang berpengaruh signifikan terhadap hasil klasifikasi. *Filter* atribut menggunakan metode *Greedy Stepwise*. Tujuh atribut dipilih untuk tujuan prediksi hasil akhir. Setelah itu proses klasifikasi untuk penentuan Model *Naive Bayes* berhasil dilaksanakan untuk prediksi hasil akhir ujian semester mahasiswa.

Dari *review* jurnal di atas, untuk membantu mencapai tujuan dalam penelitian ini maka diperoleh faktor-faktor yang dapat menjadi referensi untuk prediksi performansi mahasiswa. Adapun prediksi performansi dapat diproses dalam berbagai macam metode dengan hasil akurasi yang berbeda-beda.



### 2.7.2. Penelitian yang Terkait dengan Metode *Support Vector Machine* (SVM)

Berbagai macam penelitian terkait penggunaan *Support Vector Machine* (SVM) yang digunakan dalam berbagai bentuk klasifikasi data didapatkan oleh penulis untuk menambah referensi bagi penelitian. Metode SVM dapat diterapkan dengan dibandingkan hasilnya menggunakan metode lain, ataupun dengan perbandingan fungsi Kernel yang digunakan untuk mendapatkan hasil klasifikasi yang paling optimal. Berikut beberapa referensi yang akan dipaparkan oleh penulis guna keperluan terkait penelitian yang akan dilakukan.

Penelitian yang dilakukan oleh Siti Nurhayati, Kusriani dan Emha Taufiq Luthfi (2015) dengan judul *Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine*. Penelitian ini membahas mengenai penerapan *Support Vector Machine* yang digunakan untuk memperoleh *margin* terbaik dalam memprediksi mahasiswa *drop out*. Variabel *input* yang digunakan dalam penelitian ini yaitu data mahasiswa dan data evaluasi IP dan IPK. Jumlah kategori yang akan di prediksi memiliki 2 (dua) kategori *output* adalah mahasiswa *drop out* dan *non drop out*.

Data yang digunakan adalah data dari program studi Kesehatan masyarakat tahun 2007 dari 174 data terseleksi 15 variabel. Tahun 2008 dari 369 record data terseleksi 17 variabel. Tahun 2009 dari 229 record terseleksi 17 variabel. Tahun 2010 dari 246 record terseleksi 15 variabel. Tahun 2011 dari 237 record terseleksi 12 variabel. Pada program studi gizi tahun 2007 dari 63 record data terseleksi 11 variabel. Tahun 2008 dari 110 record data terseleksi 13 variabel. Tahun 2009 dari 85 record terseleksi 11 variabel. Tahun 2010 dari 93 record terseleksi 10 variabel. Tahun 2011 dari 37 record terseleksi 8 variabel. Proses seleksi variabel data set menggunakan metode *Principal Component Analysis* (PCA).

Prediksi *Support Vector Machine* dilakukan dengan bantuan perangkat lunak matlab. Penelitian ini menggunakan tipe kernel polynomial dan Radial basis function (RBF). Parameter nilai degree (P) 1,2,3, ..., 10, nilai cost (C) berdasarkan 0.1, 0.2, ..., 0.3 nilai gamma 0.0001, 0.001, 0.1, ..., 0.8, dan k-fold (number of validation) k=10. Setelah didapat hasil SVM pada data *training*, selanjutnya diterapkan pada data *testing*. Adapun data *testing* yang digunakan adalah data mahasiswa tahun 2011 dengan 310 record data, dan menggunakan tipe kernel polynomial, k-fold =10 nilai cost 0,5 gamma 0,001 dan degree (P) 8.

**Tabel 2.7** Hasil Pengujian *Support Vector Machine* (Nurhayati, S., dkk., 2015)

<b>Output</b>	<b>KS</b>	<b>HU</b>
Non Drop Out	283	90
Drop Out	27	20
Total Data	310	310

Keterangan:

KS : Kondisi Sebenarnya

HU : Hasil Uji

Berdasarkan hasil pengujian performa pada Tabel 2.7 dengan menggunakan data uji tahun 2011 sebanyak 310 set data dengan mahasiswa dengan mahasiswa non drop out 283 orang (KS) dan drop out 27 (KS) orang menghasilkan prediksi mahasiswa drop out 20 (HU) orang dan 290 orang (HU) non drop out. Diperoleh jumlah prediksi yang benar (sama) sebanyak 304 set data dan jumlah prediksi yang salah (tidak sama) sebanyak 6 set data. Untuk pengukuran tingkat akurasi diperoleh akurasi sebesar 98,06% dan nilai error sebesar 0.0193. Adapun Saran untuk penelitian selanjutnya perlu dilakukan antara lain melakukan analisis lebih lanjut terhadap faktor-faktor yang mempengaruhi drop out dan menggunakan data kemahasiswaan yang lebih lengkap, pemilihan parameter *Support Vector Machine* yang optimal dapat menggunakan metode *Particle Swarm Optimization (PSO)* serta menambahkan beberapa algoritma klasifikasi data mining untuk di komparasi dengan metode *Support Vector Machine*.

Penelitian yang dilakukan oleh Rizqi Agung Permana (2016) dengan judul *Seleksi Atribut pada Metode Support Vector Machine untuk Menentukan Kelulusan Mahasiswa E-Learning*. Penulis menggunakan data yang di ambil dari database platform e-learning Moodle (*Modular Object Oriented Developmental Learning Environment*) yang digunakan untuk proses belajar mengajar di Bina Sarana Informatika. Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 400 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Peneliti melakukan proses ekperimental menggunakan aplikasi *RapidMiner 5.3*.

Dari 6 atribut yang digunakan dalam seleksi atribut, yang peneliti gunakan yaitu jenis kelamin, jumlah latihan, jumlah login, jumlah akses document, jumlah forum dan jumlah pesan diperoleh bahwa jumlah latihan sebagai atribut yang paling berpengaruh

terhadap hasil kelulusan mahasiswa dengan nilai probabilitas 0.898. Hal ini membuktikan bahwa benar terjadi peningkatan akurasi ketika menggunakan algoritma *Support Vector Machine* (SVM). Hasil pengujian ini membuktikan bahwa *Support Vector Machine* (SVM) memiliki kemampuan generalisasi yang sangat baik untuk memecahkan masalah walaupun dengan sampel yang terbatas.

Penelitian yang dilakukan oleh Chinwe Peace Igiri (2015) dengan judul *Support Vector Machine–Based Prediction System for a Football Match Result*. Penelitian ini meneliti kinerja suatu multimodel SVM dengan regresi untuk prediksi hasil pertandingan sepak bola menggunakan *English Premier League* sebagai data set. Jenis kombinasi kernel Gaussian dari SVM digunakan untuk menganalisis set fitur sepakbola. Sebanyak 38 atribut yang digunakan untuk setiap pertandingan. Hasil akurasi prediksi adalah 53,3% dengan 16 set data training untuk memprediksi pertandingan diketahui hasil selama 15 pertandingan; 8 dari 15 diprediksi dengan benar, sementara 7 memiliki hasil imbang. Temuan menunjukkan bahwa SVM bukan merupakan teknik yang tepat untuk fitur set yang digunakan untuk analisis ini, karena penelitian lain dengan data sama menghasilkan akurasi sebesar 85% dengan menggunakan metode jaringan saraf tiruan. Keterbatasan penelitian ini adalah akurasi prediksi masih tergolong rendah. Penelitian lebih lanjut dapat dilakukan tentang bagaimana meningkatkan akurasi prediksi menggunakan SVM dengan penggunaan kombinasi fungsi kernel lainnya, termasuk dot, ANOVA, Epanechnikov, multiquadric, radial, dan sebagainya bisa diselidiki untuk memverifikasi kinerja fungsi tersebut dalam hal akurasi prediksi.

Penelitian yang dilakukan oleh Theopilus Bayu S, Adhistya Erna Permanasari dan Indriana Hidayah (2014) dengan judul *Komparasi Kernel pada Algoritma Support Vector Machine Studi Kasus Klasifikasi Penjurusan di SMA Saverius Sragen*. Penelitian yang dilakukan peneliti di atas membahas tentang sebuah pendekatan untuk mengklasifikasikan penjurusan pada jenjang SMA. Dataset terdiri dari 40 siswa-siswi. Setiap siswa-siswi memiliki 15 atribut. Dengan menggunakan metode pendekatan data *mining*, yaitu metode klasifikasi *Support Vector Machine* dan perangkat lunak *Rapidminer*, akan dianalisa efek dari perubahan fungsi kernel variasi faktor pinali (C) terhadap akurasi model klasifikasi.

Data masukan terdiri dari dua masukan yaitu data nilai raport kelas X dan nilai psikologi siswa. Data nilai raport meliputi nilai mata pelajaran biologi, ekonomi, fisika,

geografi, kimia, matematika, sejarah, sosiologi. Sedangkan data nilai psikologi meliputi nilai IQ, kecerdasan rasional, abstrak konseptual, analisis sintesa, logika verbal, dan logika numerik. Sedangkan keluaran yang dihasilkan adalah penjurusan SMA yaitu IPA dan IPS. Kemudian dilakukan pre-prosesing data, salah satunya dengan melakukan *scaling method* dengan menggunakan *Min-Max normalization* hingga data siap digunakan. Penggunaan min-max normalization ini didasarkan oleh penelitian yang menunjukkan bahwa metode min-max normalization menunjukkan hasil terbaik dibandingkan dengan *scaling method* yang lain. Data yang telah dilakukan pre-prosesing kemudian dibagi menjadi dua yaitu data training dan data testing dengan menggunakan teknik *k-fold cross validation*. Dataset bakat penjurusan yang telah dilakukan *k-fold cross validation* kemudian dilakukan proses training pada Sistem pelatihan SVM untuk menghasilkan model pelatihan yang terbaik.

**Tabel 2.8** Tingkat Akurasi Klasifikasi SVM (Theopilus, B. S., dkk., 2014)

Par C	TIPE KERNEL							
	Dot	Radial	Polinomial	Neural	Anova	Epacheni nikov	Gasussian Comb.	Multi- Quadratic
0.1	88.89	58.33	63.89	80.56	80.56	50.00	33.33	50.00
0.2	86.11	58.33	63.89	75.00	80.56	50.00	33.33	50.00
0.4	77.78	58.33	63.89	80.56	80.56	50.00	33.33	50.00
0.8	77.78	58.33	63.89	69.44	80.56	50.00	33.33	50.00
1.0	77.78	58.33	63.89	75.00	80.56	50.00	33.33	50.00
1.2	77.78	58.33	63.89	72.22	80.56	50.00	33.33	50.00
1.4	77.78	58.33	63.89	72.22	80.56	50.00	33.33	50.00
1.8	77.78	58.33	63.89	66.67	80.56	50.00	33.33	50.00

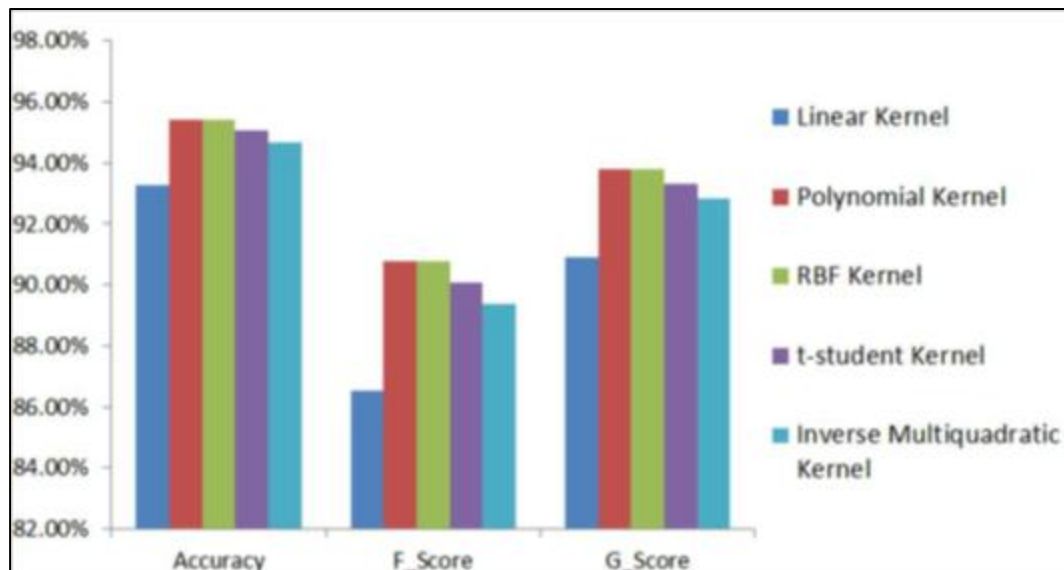
Pada Tabel 2.8 memperlihatkan bahwa memperbesar nilai parameter C tidak memperbaiki tingkat akurasi yang diberikan. Tiap kernel memiliki tingkat akurasi yang rendah, hal ini disebabkan oleh keterbatasan data pelatihan yang digunakan. Tingkat akurasi terbaik adalah 88.89% diperoleh dari tipe kernel dot dengan nilai parameter C sebesar 0.1. Pengembangan penelitian yang akan datang perlu dilakukan dengan dataset yang relatif besar agar tingkat keakuratan performansi model terjamin. Perlu dilakukan mengenai metode pemilihan atribut, dan penentuan parameter SVM yang cocok.

Penelitian yang dilakukan oleh Krupal S. Parikh dan Trupti P. Shah (2016) yang berjudul *Support Vector Machine – a Large Margin Classifier to Diagnose Skin Illnesses*. Penelitian ini dipublikasikan pada *3rd International Conference on Innovations in Automation and Mechatronics Engineering*, ICIAME 2016. Penelitian yang dilakukan membahas tentang analisis model diagnostik untuk mengklasifikasikan penyakit kulit ke dalam algoritma SVM. Dalam penelitian ini, penulis menggunakan algoritma *one-to-one* untuk data *multi-class* yang akan digunakan. *Database* diperoleh dari Departemen Kulit & V.D., Rumah Sakit Shrikrishna, Karamsad, Gujarat, India.

Dalam penelitian ini, dilakukan klasifikasi beberapa penyakit kulit yang umum yaitu, infeksi bakteri, infeksi jamur, Kudis dan Eksim. Data yang didapat berupa performa rinci di bawah bimbingan dokter kulit terkemuka dengan mendiagnosa 470 pasien. Data atribut bagi setiap pasien didapatkan dari wawancara para dokter. Ada 47 fitur dan 470 kasus. Dari 470 kasus 139 kasus yang untuk penyakit infeksi bakteri, 146 untuk infeksi jamur, 98 untuk Eksim dan 87 untuk Kudis. Penelitian ini berfokus pada perbedaan hasil setiap penggunaan fungsi Kernel yang diterapkan di SVM untuk diagnosa medis. Hasil analisis klasifikasi tiap fungsi Kernel ditampilkan pada Tabel 2.9 dan grafik Gambar 2.7 berikut.

**Tabel 2.9** Performansi Fungsi Kernel SVM dengan parameter C=4096  
(Parikh, K. S. dan Shah, T. P., 2016)

Kernel Function $K(x,y)$	Values of kernel Parameters	Accuracy	F-Score	G-Score
<b>Linear:</b> $x^T y + c$	$c = 10$	93.26%	86.52%	90.91%
<b>Polynomial:</b> $(\alpha x^T + c)^d$	$\alpha = 2, c = 10,$ $d = 3$	95.39%	90.78%	93.80%
<b>Radial Basis Function (RBF):</b> $exp(-\gamma \ x - y\ ^2)$	$\gamma = 0.1$	95.39%	90.78%	93.80%
<b>t-Student:</b> $1/(1 + \ x - y\ ^d)$	$d = 2$	95.04%	90.07%	93.32%
<b>Inverse Multiquadratic</b> $1/(c^2 + \ x - y\ ^2)$	$c = 10$	94.68%	89.36%	92.84%



**Gambar 2.7** Grafik Hasil Akurasi, F-Score dan G-Score untuk Macam-macam Kernel (Parikh, K. S. dan Shah, T. P., 2016)

Hasil perhitungan menunjukkan bahwa fungsi Kernel *Radial Basis Function* (RBF) dan kernel polinomial memberikan akurasi yang lebih baik dari pada kernel lain. Hal ini juga menunjukkan bahwa metode Kernel sangat efisien jika diterapkan dalam cara yang kreatif dan dapat memecahkan berbagai masalah dalam sains dan teknik. Untuk saran ke depan, fungsi kernel baru yang lebih kuat dapat diusulkan untuk meningkatkan akurasi *classifier* sehingga hasil diagnosis bisa dilakukan lebih akurat.

## 2.8. Posisi Penelitian

Tujuan dari review jurnal dan tinjauan pustaka dan terhadap penelitian terdahulu adalah agar penulis dapat mengetahui posisi penelitian saat ini. Pada penelitian ini yang dijadikan rujukan adalah faktor-faktor yang mempengaruhi prediksi performansi mahasiswa dan metodologi yang digunakan. Adapun lingkup penelitian dan metodologi yang digunakan pada penelitian terdahulu seperti yang terlihat pada Tabel 2.10.

**Tabel 2.10** Lingkup dan Metode yang Digunakan pada Penelitian Terdahulu

No.	Lingkup Penelitian	Metode yang digunakan	Penulis
1.	Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector	Multi Layer Perception, Naive Bayes Classifier, SMO, J48 dan REPTree dengan perangkat lunak WEKA	Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan (2015)
2.	Predicting Students Final GPA Using Decision Trees: A Case Study	<i>Decision Tree</i> J48	Mashael A. Al-Barrak dan Muna Al-Razgan (2016)
3.	Of Students Academic Performance Rates Using Artificial Neural Networks (ANNs)	Artificial Neural Networks	O. C. Asogwa dan A. V. Oladugba (2015)
4.	A Comparison of Logistic Regression Model and Artificial Neural Networks in Predicting of Student's Academic Failure	Regresi Logistik dan Artificial Neural Networks	Saeed Hosseini Teshnizi, dkk (2015)
5.	Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier	Naïve bayes Classifier	Humera Shaziya, dkk (2015)
6.	Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine	SVM dengan Kernel Polynomial dan Radial basis function (RBF)	Siti Nurhayati, Kusri dan Emha Taufiq Luthfi (2015)

No.	Lingkup Penelitian	Metode yang digunakan	Penulis
7.	Seleksi Atribut pada Metode Support Vector Machine untuk Menentukan Kelulusan Mahasiswa E-Learning	SVM dengan RapidMiner 5.3	Rizqi Agung Permana (2016)
8.	Support Vector Machine–Based Prediction System for a Football Match Result	SVM dengan Kernel Gaussian	Chinwe Peace Igiri (2015)
9.	Komparasi Kernel pada Algoritma <i>Support Vector Machine</i> Studi Kasus Klasifikasi Penjurusan di SMA Saverius Sragen	SVM dengan Kernel Dot Radial, Polinomial, Neural, Anova, Epacheninikov, Gasussian Comb., Multi-Quadratic	Theopilus Bayu S, Adhistya Erna Permanasari dan Indriana Hidayah (2014)
10.	Support Vector Machine – a Large Margin Classifier to Diagnose Skin Illnesses	SVM dengan Kernel Linear, Polynomial, Radial Basis Function (RBF), t-Student, Inverse Multiquadratic	Krupal S. Parikh dan Trupti P. Shah (2016)

Pada Tabel 2.10 nomor 1 s.d 5 merupakan penelitian prediksi performansi yang menggunakan berbagai macam metode klasifikasi dan referensi faktor-faktor penentu prediksi yang sudah diolah. Adapun nomor 6 s.d 10 merupakan penelitian penerapan *Support Vector Machine* dengan berbagai macam kasus klasifikasi. Penerapan SVM dapat digunakan untuk mengkomparasi hasilnya dengan metode yang lain ataupun mengkomparasi fungsi Kernel yang digunakan.



Berdasarkan Tabel 2.10, yang membedakan dan memposisikan penelitian ini terhadap penelitian terdahulu diantaranya:

1. Menggunakan metode klasifikasi *Support Vector Machine* dengan melakukan komparasi berbagai macam fungsi Kernel untuk mendapatkan model SVM dengan hasil akurasi terbaik.
2. Objek penelitian dilakukan untuk klasifikasi prediksi kinerja mahasiswa sebagai data *training* dan data *testing*, yang kemudian nantinya dapat diaplikasikan sebagai salah satu pendukung kebijakan pengelola program studi dalam penerimaan mahasiswa baru.

*halaman ini sengaja dikosongkan*

## **BAB 3**

### **METODOLOGI PENELITIAN**

Metode penelitian adalah langkah dan prosedur yang akan dilakukan untuk mencapai tujuan dan mendapatkan jawaban atas permasalahan dalam penelitian. Langkah dan prosedur ini merupakan perwujudan dari kerangka pikir penelitian.

#### **3.1. Jenis Penelitian**

Secara teoritis, penelitian ini termasuk dalam jenis penelitian tindakan (*action research*) yaitu jenis penelitian yang bertujuan untuk mengembangkan keterangan baru untuk mengatasi masalah-masalah yang ada dalam dunia kerja atau kebutuhan praktis manusia lainnya. Pendekatan penelitian ini merupakan pendekatan yang bertujuan untuk mengembangkan metode kerja yang paling efisien, sehingga biaya produksi dapat ditekan dan produktivitas lembaga atau institusi dapat meningkat. Penelitian tindakan ditujukan untuk memberikan andil pada pemecahan masalah praktis dalam situasi problematik yang mendesak dan pada pencapaian tujuan ilmu sosial melalui kolaborasi patungan dalam rangka kerja etis yang saling berterima (Rapoport, 2006). Proses penelitian bersifat dari waktu ke waktu, antara “*finding*” pada saat penelitian, dan “*action learning*”. Dengan demikian *action research* merupakan penelitian yang menghubungkan antara teori yang sudah ada dengan praktik langsung di lapangan.

Penelitian ini menghubungkan teori algoritma klasifikasi *Support Vector Machine* dengan praktik untuk diterapkan dalam klasifikasi prediksi kinerja mahasiswa yang ada di suatu program studi perguruan tinggi. Penulis bermaksud untuk mencari metode kerja yang paling efisien untuk meningkatkan produktivitas pada institusi. Pencarian model terbaik dari berbagai macam fungsi Kernel dalam SVM diharapkan dapat memaksimalkan hasil akurasi pada prediksi kinerja akademik tersebut. Dengan diketahuinya prediksi kinerja mahasiswa di awal waktu, maka pengelola program studi bisa mendapatkan kebermanfaatan dari penelitian untuk keputusan atau tindakan evaluasi ke depan bagi mahasiswa yang bersangkutan.

Penggolongan jenis penelitian menurut pendekatan analitik, dibagi menjadi 2 (dua) macam yaitu penelitian kualitatif dan penelitian kuantitatif. Penelitian ini

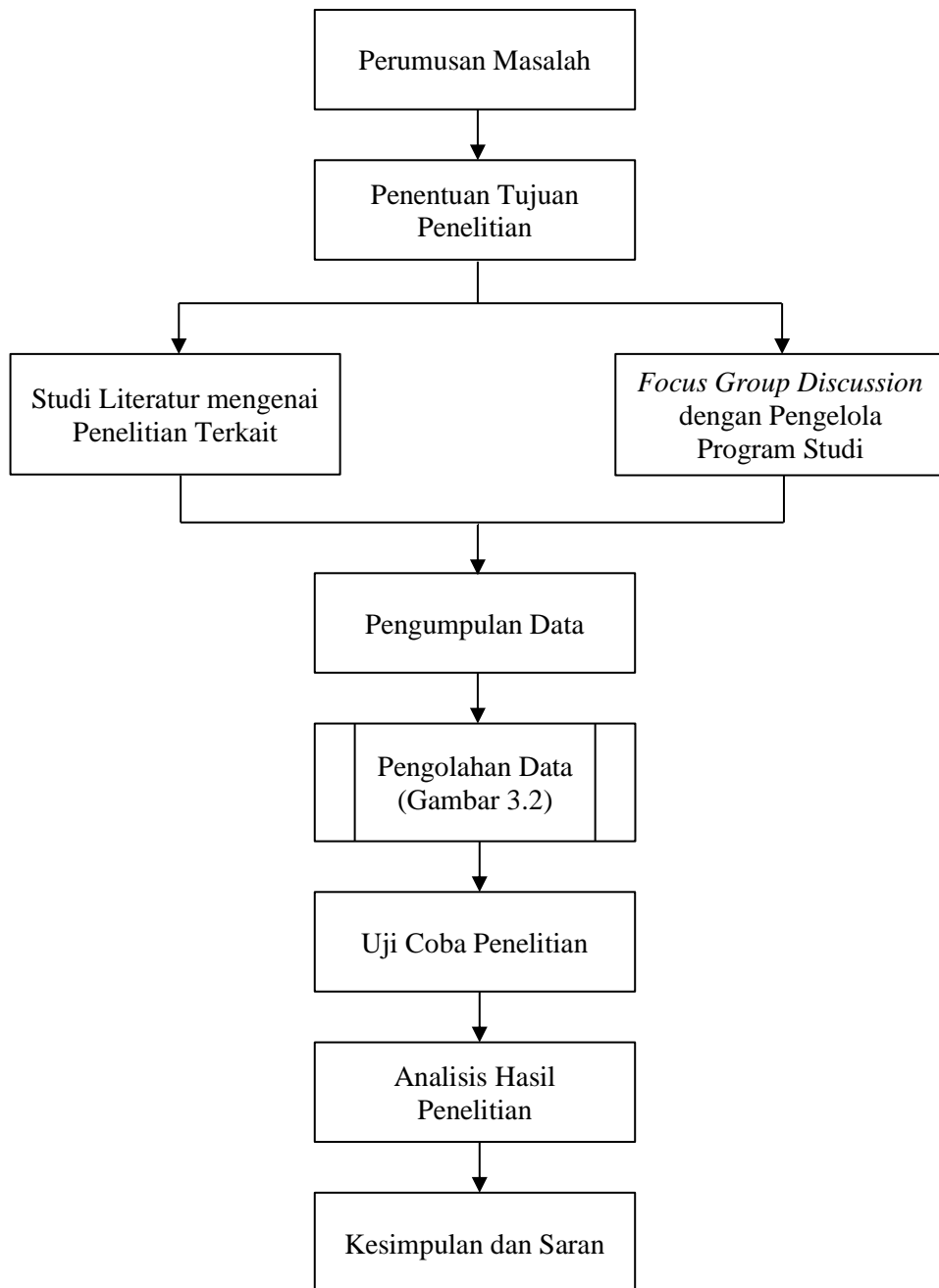
menggunakan pendekatan kuantitatif yang menekankan analisisnya pada data-data numerikal (angka-angka) yang diolah dengan metode statistik. Pada dasarnya pendekatan kuantitatif dilakukan pada jenis penelitian inferensial dan menyandarkan kesimpulan hasil penelitian pada suatu probabilitas kesalahan penolakan hipotesis nihil. Dengan metoda kuantitatif akan diperoleh signifikansi perbedaan kelompok atau signifikansi hubungan antar variabel yang diteliti. Pada umumnya, penelitian kuantitatif merupakan penelitian dengan jumlah sampel besar.

Data-data yang dibutuhkan dalam penelitian ini awalnya bersifat kategorikal kemudian nantinya dikonversi menjadi data numerikal. Data yang sudah dilakukan proses konversi tersebut kemudian dapat dengan mudah diolah dengan metode *Support Vector Machine* untuk menghasilkan prediksi klasifikasi kinerja mahasiswa.

### **3.2. Diagram Alir Penelitian**

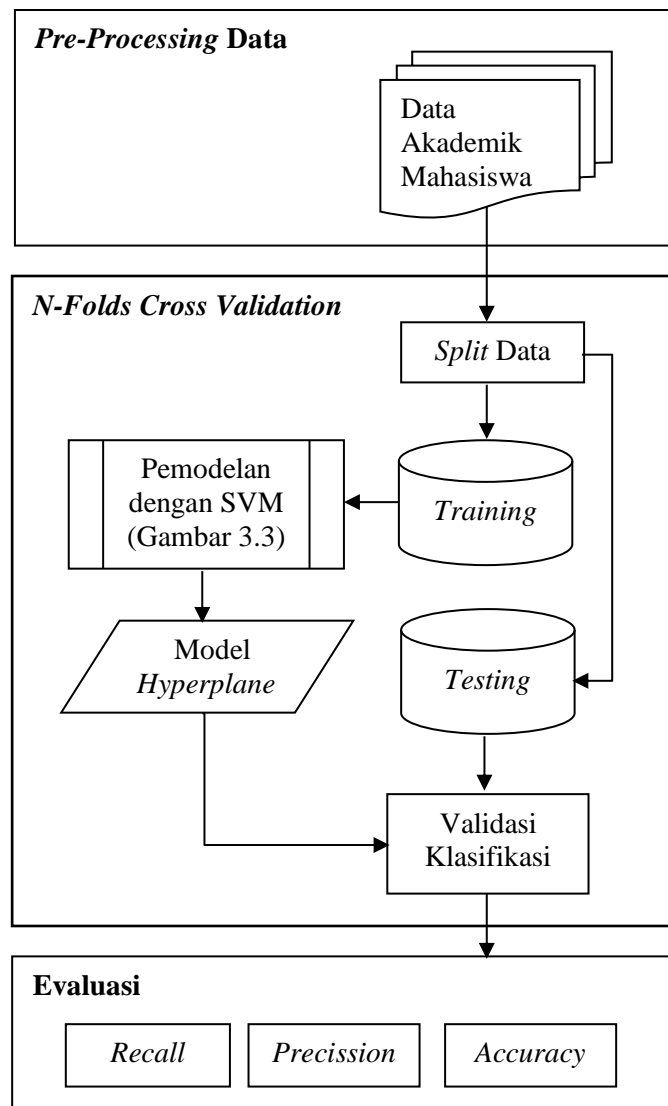
Untuk mencapai tujuan penelitian diperlukan perencanaan tahapan kegiatan yang akan dilakukan. Tahapan ini yang nantinya dijadikan sebagai pedoman pengerjaan penelitian hingga terselesaikannya penelitian ini dengan baik. Berikut adalah proses secara general kerangka tahapan penelitian Prediksi Kinerja Mahasiswa menggunakan *Support Vector Machine* untuk Pengelola Program Studi di Perguruan Tinggi yang dapat dilihat pada Gambar 3.1.

Gambar 3.1 menunjukkan diagram penelitian diawali dengan perumusan masalah, seperti yang sudah dijelaskan sebelumnya. Kemudian dilanjutkan dengan penetapan tujuan penelitian yaitu untuk membuat model prediksi kinerja mahasiswa yang nantinya dapat digunakan oleh pengelola program studi dengan menggunakan metode *Support Vector Machine*. Tahap selanjutnya penulis melakukan proses studi literatur mengenai kepustakaan terkait dengan penelitian. Selain mencari literatur tersebut, penulis juga melakukan *Focus Group Discussion* (FGD) dengan pengelola program studi untuk mempertimbangkan faktor-faktor yang mempengaruhi kinerja mahasiswa dan *output* prediksi yang diperlukan. Kemudian dilanjutkan dengan tahapan proses pengumpulan data. Data yang sudah terkumpul sesuai kebutuhan disimpan dan diproses dalam tahap pengolahan data. Tahapan selanjutnya adalah uji coba sesuai skenario yang dirancang serta penarikan kesimpulan dan saran untuk penelitian.



**Gambar 3.1** Kerangka Tahapan Penelitian

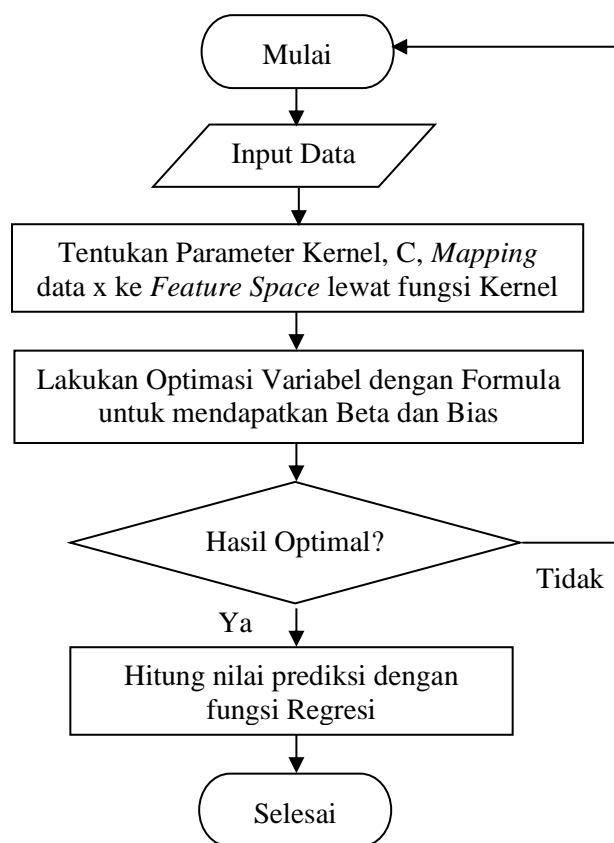
Adapun proses pengolahan data secara detail ditunjukkan pada Gambar 3.2. Bentuk klasifikasi prediksi kinerja mahasiswa dalam penelitian ini dilakukan dengan teknik *supervised learning*, yakni teknik pendekatan data *mining* dengan membuat model yang mengacu pada data *set* histori sebelumnya untuk dilihat pola data sehingga bisa diimplementasikan pada data uji.



**Gambar 3.2** Diagram Proses Pengolahan data

Pengolahan data diawali dengan tahapan *Pre-processing* data, yakni suatu langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas dengan input yang baik untuk data *mining tools*. Jika data masukan tidak berkualitas, maka hasil data *mining* juga tidak akan berkualitas. Dalam tahap ini, akan dilakukan analisis distribusi data guna melihat *plot* distribusi data pada variabel *input* maupun *output*. Tahapan ini juga merupakan tahap untuk melakukan normalisasi pada data. Data mentah yang didapatkan dikonversi menjadi bentuk data yang dibutuhkan sehingga siap untuk diolah. Tahapan ini meliputi proses seleksi data yang akan digunakan sebagai variabel masukan untuk proses klasifikasi dengan uji korelasi pada data.

Tahap kedua adalah tahap pemodelan data menggunakan implementasi algoritma *Support Vector Machine* (SVM) dan teknik *n-Folds Cross Validation*. Teknik *n-Folds Cross Validation* ini adalah metode validasi silang. Metode ini membagi data menjadi dua bagian, yaitu data pelatihan (*training*) dan data pengujian (*testing*). Selanjutnya, setelah data diuji dilakukan proses silang dimana data pengujian lantas dijadikan data pelatihan ataupun sebaliknya, data pelatihan sebelumnya dijadikan kini menjadi data pengujian. Data *training* yang sudah ada siap dilakukan proses pemodelan klasifikasi menggunakan SVM hingga mendapatkan model *hyperplane*. Adapun proses detail diagram alir algoritma SVM ditampilkan pada Gambar 3.3. Setelah model SVM untuk klasifikasi didapatkan, model *hyperplane* tersebut diimplementasikan pada data *testing*. Tahap ketiga adalah proses evaluasi dengan menggunakan pengukuran evaluasi klasifikasi yakni pengukuran *Recall*, *Precision* dan Akurasi.



**Gambar 3.3** Flowchart Algoritma SVM

### **3.3. Pengumpulan Data**

Tujuan pengumpulan data yang diperlukan dari penelitian ini adalah untuk mendapatkan faktor-faktor yang mempengaruhi performansi mahasiswa guna keberhasilan penelitian. Pada sub bagian Pengumpulan Data, berikut akan dipaparkan mengenai populasi dan sampel penelitian, data primer dan data sekunder yang dibutuhkan, serta definisi operasional data yang dibutuhkan.

#### **3.3.1. Populasi dan Sampel Penelitian**

Populasi adalah keseluruhan subyek penelitian. Apabila seseorang ingin meneliti semua elemen yang ada dalam wilayah penelitian, maka penelitiannya merupakan penelitian populasi atau studi populasi atau studi sensus (Sabar, 2007). Sedangkan menurut Sugiyono pengertian populasi adalah wilayah generalisasi yang terdiri atas: obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Sugiyono, 2011). Sehingga dapat disimpulkan bahwa populasi bukan hanya orang tapi juga obyek dan benda-benda alam yang lain. Populasi bukan sekedar jumlah yang ada pada obyek/subyek yang dipelajari, tetapi meliputi karakteristik/sifat yang dimiliki oleh subyek atau obyek itu.

Pengertian dari sampel adalah sebagian dari subyek dalam populasi yang diteliti, yang sudah tentu mampu secara representatif dapat mewakili populasinya (Sabar, 2007). Menurut Sugiyono, sampel adalah bagian atau jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Bila populasi besar, dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, misal karena keterbatasan dana, tenaga dan waktu, maka peneliti akan mengambil sampel dari populasi itu. Apa yang dipelajari dari sampel itu, kesimpulannya akan diberlakukan untuk populasi. Untuk itu sampel yang diambil dari populasi harus betul-betul representatif (Sugiyono,2011).

Populasi dalam penelitian ini adalah Mahasiswa Program Studi Magister Statistika ITS yang akan diprediksi performansi mahasiswanya. Adapun sampel data yang digunakan adalah data demografi maupun data akademik mahasiswa untuk angkatan 2011 hingga 2015. Pertimbangan penggunaan sampel tersebut adalah bahwa setiap data akademik yang disimpan dalam kearsipan adalah data akademik pada lima



tahun terakhir, oleh karena itu data yang digunakan terbatas pada data mahasiswa 2011 hingga 2015.

### **3.3.2. Data Primer**

Data primer dalam penelitian kali ini diperoleh dengan cara melakukan observasi langsung dan *focus group discussion* (FGD). Wawancara yang dilaksanakan dengan narasumber yang berkompeten (pakar)/*expert* dalam penelitian ini adalah untuk mendefinisikan faktor-faktor yang mempengaruhi hasil performansi mahasiswa yang nantinya akan diprediksi dalam penelitian ini. Adapun data faktor-faktor yang mempengaruhi hasil performansi mahasiswa Program Studi Magister Statistika antara lain Jenis Kelamin, Jalur Masuk S2, Asal Perguruan Tinggi S1, Jurusan S1, Indeks Prestasi Kumulatif S1, Lama Studi S1, Waktu Tunggu, Nilai Toefl, Nilai Tes Potensi Akademik, Asal Daerah, Usia saat pendaftaran S2, Status Keikutsertaan Matrikulasi dan Status Bekerja.

### **3.3.3. Data Sekunder**

Data sekunder adalah data yang tidak langsung memberikan informasi kepada peneliti, misalnya penelitian harus melalui orang lain atau mencari melalui dokumen (Sugiyono, 2011). Data sekunder adalah data hasil pengamatan dari pihak-pihak atau instansi terkait yang dapat menunjang variabel-variabel penelitian yang dilakukan. Data sekunder yang digunakan dalam penelitian ini meliputi data faktor-faktor performansi yang sudah dijelaskan sebelumnya yang ada untuk setiap mahasiswa. Beberapa data tersebut sudah ada dalam SIM Akademik perguruan tinggi, namun beberapa data masih harus dimasukkan secara manual untuk mendapatkannya.

### **3.3.4. Definisi Operasional Variabel**

Definisi operasional variabel digunakan untuk mempermudah peneliti dalam pengumpulan data, baik data primer maupun data sekunder pada objek penelitian ini. Berikut akan dijelaskan secara rinci mengenai variabel yang digunakan dalam penelitian termasuk indikator pengukuran serta definisi operasionalnya. Adapun yang dimaksud definisi operasional variabel dalam pembahasan ini adalah proses mendefinisikan variabel dengan tegas, sehingga menjadi faktor-faktor yang dapat diukur. Penentuan faktor-faktor definisi variabel dan indikatornya ditetapkan berdasarkan hasil *Focus*

*Group Discussion* yang dilakukan oleh peneliti bersama pakar, dalam hal ini adalah Manajemen Prodi Magister Statistika ITS. Variabel dan definisi operasional variabel pada penelitian ini dituangkan pada Tabel 3.1.

**Tabel 3.1** Variabel dan Definisi Operasional Variabel

<b>No.</b>	<b>Variabel</b>	<b>Definisi Operasional</b>	<b>Indikator</b>
1.	Jenis Kelamin	Perbedaan sistematik pada setiap individu yang dibagi dalam bentuk Laki-laki maupun Perempuan	Laki-laki atau Perempuan
2.	Asal Daerah	Daerah tempat tinggal calon mahasiswa sebelum menempuh studi	Jarak tempuh daerah asal dengan institusi
3.	Usia	Usia saat calon mahasiswa saat akan menempuh jenjang Prodi Magister (S2)	Usia dalam skala Tahun
4.	Status Bekerja	Status pekerjaan calon mahasiswa saat akan menempuh jenjang Prodi Magister (S2)	Bekerja atau tidak
5.	Jalur Masuk S2	Bagian jalan atau cara yang diperuntukkan bagi calon mahasiswa yang ingin masuk ke jenjang prodi Magister (S2)	Jenis jalur masuk yang terdiri dari Biaya Sendiri, Beasiswa ataupun Kerjasama Instansi
6.	Asal Perguruan Tinggi S1	Perguruan Tinggi yang telah ditempuh oleh calon mahasiswa pada tahap sebelumnya, yakni jenjang Sarjana	Akreditasi Perguruan Tinggi
7.	Asal Jurusan S1	Jurusan atau bidang minat yang telah ditempuh oleh calon mahasiswa pada tahap sebelumnya, yakni jenjang Sarjana	Kelinieritasan Jurusan atau Prodi
8.	Indeks Prestasi Kumulatif S1	Nilai kumulatif dari calon mahasiswa selama menempuh jenjang Sarjana	Nilai IPK dengan Range Nilai 0-4
9.	Lama Studi S1	Waktu yang dibutuhkan calon mahasiswa untuk menyelesaikan studi pada jenjang Sarjana	Jumlah Semester Tempuh

No.	Variabel	Definisi Operasional	Indikator
10.	Waktu Tunggu	Waktu yang dibutuhkan calon mahasiswa untuk menunggu dari kelulusannya pada jenjang Sarjana hingga memutuskan untuk mendaftar di jenjang Magister (S2)	Jumlah Semester Tempuh
11.	Nilai Toefl	<i>Test of English as a Foreign Language</i> , yakni ujian kemampuan penggunaan bahasa Inggris yang diakui oleh institusi	Nilai Toefl dengan Range 310-677
12.	Nilai Tes Potensi Akademik	Sebuah tes yang bertujuan untuk mengetahui bakat dan kemampuan calon mahasiswa di bidang keilmuan atau akademis yang diakui oleh institusi	Nilai TPA dengan Range 200-800
13.	Status S2	Keadaan atau kondisi terakhir yang disandang sebagai mahasiswa saat menempuh jenjang Magister	Lulus, Mengundurkan Diri, <i>Drop Out</i>
14.	Lama Studi S2	Waktu yang dibutuhkan calon mahasiswa untuk menyelesaikan studi pada jenjang Magister	Jumlah semester tempuh
15.	Indeks Prestasi Kumulatif S2	Nilai kumulatif dari mahasiswa selama menempuh jenjang Magister	Nilai IPK dengan Range Nilai 0-4

### 3.4. Identifikasi Variabel

Identifikasi variabel penelitian dilakukan melalui proses studi literatur pada penelitian sebelumnya sebagai dasar yang akan digunakan pada survei pendahuluan guna memperoleh variabel yang relevan untuk dimodelkan dan dianalisa. Berikut akan dilakukan identifikasi variabel yang diperlukan dalam penelitian. Variabel dalam penelitian ini terdiri dari dua jenis yakni Variabel Masukan dan Variabel Pengeluaran.

### 3.4.1. Variabel Masukan (*Input*)

Variabel Masukan didefinisikan sebagai variabel yang diberikan pada sistem untuk menghasilkan respon tertentu. Variabel masukan (*input*) juga disebut sebagai variabel independen. Variabel independen adalah tipe variabel yang menjelaskan atau mempengaruhi variabel yang lain. Dalam penelitian ini, variabel masukan yang dimaksud adalah data-data mahasiswa yang diharapkan dapat mempengaruhi hasil prediksi performansi mahasiswa dalam bentuk klasifikasi. Variabel yang sudah didefinisikan sebelumnya pada Tabel 3.1 kemudian dilakukan konversi menjadi data numerik supaya dapat diolah untuk kebutuhan penelitian ini. Beberapa variabel merupakan variabel yang sudah bersifat numerik sehingga tidak perlu dilakukan konversi data. Adapun variabel yang tidak perlu dilakukan konversi data disajikan pada Tabel 3.2. Variabel-variabel tersebut merupakan jenis variabel yang kuantitatif dan memiliki besaran angka yang jelas.

**Tabel 3.2** Variabel Masukan Tanpa Proses Konversi

No	Nama Variabel
1.	Usia
2.	Indeks Prestasi Kumulatif S1
3.	Lama Studi S1
4.	Waktu Tunggu
5.	Nilai Toefl
6.	Nilai TPA

Adapun variabel yang perlu dilakukan proses konversi sebagai variabel masukan antara lain Jenis Kelamin, Asal Kota, Status Bekerja, Jalur Masuk S2, Asal Perguruan Tinggi S1, dan Asal Jurusan S1. Keenam variabel tersebut perlu dikelompokkan secara kategorikal terlebih dahulu kemudian dikonversi menjadi data numerik. Tabel 3.3 merupakan daftar variabel masukan dengan proses konversi.

**Tabel 3.3** Variabel Masukan dengan Proses Konversi

No	Nama Variabel	Konversi Data Numerik							
		1	2	3	4	5	6	7	8
1.	Jenis Kelamin	Laki-laki	Perempuan	-	-	-	-	-	-
2.	Asal Daerah	Jawa Timur	Luar Jawa Timur	-	-	-	-	-	-
3.	Status Bekerja	Sudah	Belum	-	-	-	-	-	-
4.	Jalur Masuk S2	Biaya Sendiri	Joint Degree	Fresh Graduate	BPPS	BPPDN, Dikti	LPDP	Fast Track	Kerjasama Instansi
5.	Asal PT S1	Akreditasi A	Akreditasi B	Akreditasi C	Belum Akreditasi	-	-	-	-
6.	Asal Jurusan S1	Statistika	Matematika	Pendidikan Matematika	Lainnya	-	-	-	-

Pada Tabel 3.3 menunjukkan terdapat tujuh variabel masukan yang perlu dilakukan proses konversi menjadi data numerik. Variabel Jenis Kelamin yang memiliki data berupa Laki-laki dan Perempuan dikonversi menjadi data numerik 1 untuk laki-laki dan 2 untuk perempuan. Variabel Asal Daerah dikategorikan menjadi 2 (dua) kategori yakni daerah Jawa Timur dan daerah luar Jawa Timur. Variabel Status Bekerja dikonversikan menjadi dua nilai yakni 1 untuk Sudah bekerja dan 2 untuk Belum bekerja. Adapun Variabel Jalur Masuk dikategorikan menjadi 8 (delapan) kategori yakni Biaya Sendiri, *Joint Degree*, *Fresh Graduate*, BPPS, BPPDN dan Dikti, LPDP, *Fast Track* dan Jalur Kerjasama Instansi. Selanjutnya Variabel Asal Perguruan Tinggi S1 dikategorikan menurut hasil akreditasi institusi, sehingga nilai 1 untuk PT yang memiliki Akreditasi A, nilai 2 untuk PT yang memiliki Akreditasi B, nilai 3 untuk PT yang memiliki Akreditasi C dan nilai 4 untuk PT yang belum terakreditasi oleh BAN-PT. Variabel Asal Jurusan S1 dikategorikan sesuai dengan kelinieritasan program studi. Untuk mahasiswa yang berasal dari Program Studi Statistika memiliki nilai 1, nilai 2 untuk Program Studi Matematika, nilai 3 untuk Program Studi Pendidikan Matematika, nilai 4 untuk mahasiswa yang memiliki background studi selain yang sudah disebutkan sebelumnya.

### 3.4.2. Variabel Keluaran (*Output*)

Variabel Keluaran didefinisikan sebagai variabel yang dihasilkan oleh sistem dari pengolahan variabel masukan sebelumnya. Variabel keluaran (*output*) juga disebut sebagai variabel dependen. Besarnya perubahan pada variabel ini tergantung dari besaran variabel masukan. Setiap kali terjadi perubahan sekian satuan pada variabel masukan, maka diharapkan akan mengakibatkan perubahan variabel keluaran sekian satuan juga.

Dalam penelitian ini, variabel keluaran yang dimaksud adalah data level kinerja dan keberhasilan mahasiswa. Kinerja performansi dalam hal ini merupakan tolak ukur keberhasilan seorang mahasiswa dalam menempuh studi di perguruan tinggi. Berdasarkan referensi dari Buku Panduan Akademik ITS Tahun 2014 dan wawancara dengan pakar dalam hal ini adalah pengelola program studi yang ada di Program Studi Magister Statistika, level kinerja performansi mahasiswa dapat dipengaruhi oleh beberapa faktor. Skala level kinerja mahasiswa sebagai *output* klasifikasi yang akan dihasilkan dalam penelitian ini, tersaji dalam Tabel 3.4.

**Tabel 3.4** Variabel Keluaran Level Kinerja Mahasiswa

<b>Level Kinerja Mahasiswa</b>	<b>Status S2</b>	<b>Lama Studi S2</b>	<b>IPK S2</b>
Level 7	L	< 4 Semester	> 3,75
Level 6	L	4 Semester	> 3,75
Level 5	L	< 4 semester	3 – 3,75
Level 4	L	4 semester	≤ 3,75
Level 3	L	> 4 Semester	> 3,75
Level 2	L	> 4 Semester	3 – 3,75
Level 1	L	> 4 Semester	< 3
Level 0	DM/DO		

Keterangan :

L : Lulus

DM : Mengundurkan Diri

DO : *Drop Out*

Pada Tabel 3.4 menunjukkan bahwa level kinerja dan keberhasilan mahasiswa dipengaruhi oleh tiga faktor utama antara lain Status Lulus S2, Lama Studi S2 dan Indeks Prestasi Kumulatif (IPK) S2. Status lulus yang dimaksudkan adalah keterangan mahasiswa tersebut yang masuk dalam kategori Lulus, Mengundurkan diri, ataupun *Drop Out*. Klasifikasi level performansi inilah yang nantinya akan menjadi variabel

keluaran untuk memprediksi Performansi Mahasiswa dalam penelitian ini. Level kinerja tersebut dikategorikan berdasarkan pertimbangan ketepatan waktu studi mahasiswa dan besaran IPK mahasiswa dengan beberapa asumsi, antara lain:

1. Level kinerja paling tinggi dikategorikan bagi mahasiswa yang memiliki IPK *cumlaude*, yakni lebih dari 3,75 dengan waktu masa studi yang lebih singkat.
2. Mahasiswa yang dikategorikan memiliki IPK lebih dari 3,75 dalam kurun waktu tepat 4 semester (Level 6) diasumsikan lebih baik dibandingkan dengan mahasiswa yang menempuh perkuliahan kurang dari 4 semester namun memiliki IPK tidak lebih dari 3.75 (Level 5).
3. Mahasiswa dengan IPK kurang dari 3,75 dengan masa studi tepat 4 semester (Level 4) diasumsikan lebih baik dibandingkan dengan mahasiswa yang memiliki IPK lebih dari 3,75 dengan masa studi yang lebih dari 4 semester (Level 3).
4. Mahasiswa yang memiliki status akhir *Drop Out* ataupun mengundurkan diri diasumsikan memiliki level kinerja yang paling rendah tanpa melihat IPK maupun lama studi mahasiswa.

*halaman ini sengaja dikosongkan*



## BAB 4

### HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil dan pembahasan penelitian. Sebagai alat bantu pengolahan data, penulis menggunakan *software* Matlab R2011b. Adapun hasil dan pembahasan dibagi dalam beberapa subbab yakni Distribusi Data, Implementasi *Support Vector Machine* pada Matlab, Pelatihan dan Pengujian, serta Analisis penggunaan metode SVM.

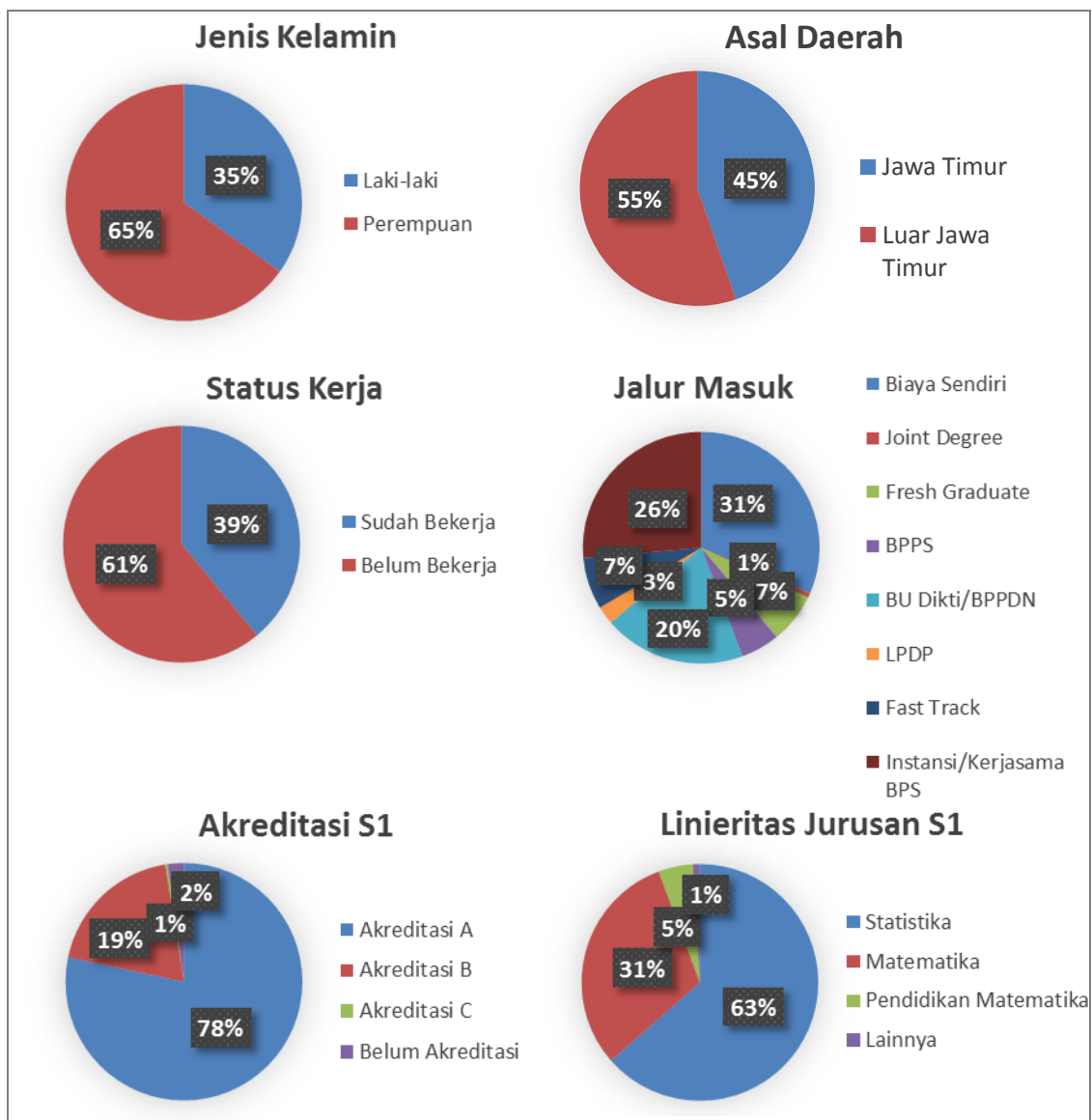
#### 4.1. Distribusi Data

Tujuan distribusi data yang akan disajikan dalam bagian ini adalah untuk memberikan gambaran secara umum mengenai penyebaran data penelitian. Bagian ini memuat prosentase penyebaran data pada setiap variabel masukan (*input*) maupun variabel keluaran (*output*) dalam penelitian. Adapun akan dijelaskan distribusi data untuk variabel berjenis kategorikal. Variabel masukan (*input*) yang berjenis kategorikal dan penyebaran distribusi datanya ditampilkan pada Tabel 4.1 dan Gambar 4.1.

**Tabel 4.1** Distribusi Data Variabel Masukan

VARIABEL MASUKAN		JUMLAH	PROSENTASE
GENDER	Laki-laki	111	34.9%
	Perempuan	207	65.1%
ASAL DAERAH	Jawa Limur	142	44.7%
	Luar Jawa Timur	176	55.3%
STATUS KERJA	Sudah Bekerja	124	39.0%
	Belum Bekerja	194	61.0%
JALUR MASUK	Biaya Sendiri	100	31.4%
	Joint Degree	2	0.6%
	Fresh Graduate	22	6.9%
	BPPS	17	5.3%
	BU Dikti/BPPDN	62	19.5%
	LPDP	8	2.5%
	Fast Track	23	7.2%
AKREDITASI S1	Instansi/Kerjasama BPS	84	26.4%
	Akreditasi A	249	78.3%
	Akreditasi B	61	19.2%
	Akreditasi C	1	0.3%

VARIABEL MASUKAN		JUMLAH	PROSENTASE
LINIERITAS JURUSAN S1	Belum Akreditasi	7	2.2%
	Statistika	202	63.5%
	Matematika	98	30.8%
	Pendidikan Matematika	15	4.7%
	Lainnya	3	0.9%
<b>Total</b>		<b>318</b>	<b>100.0%</b>



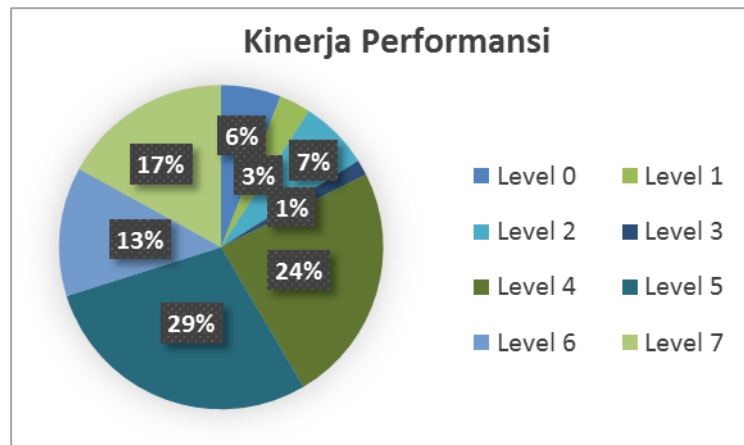
**Gambar 4.1** Grafik Distribusi Data Variabel Masukan

Tabel 4.1 dan Gambar 4.1 memperlihatkan distribusi data pada setiap kategori yang ada di variabel masukan (*input*). Pada variabel Jenis Kelamin, terdapat 65% data mahasiswa didominasi oleh kategori Perempuan dan 35% sisanya adalah data mahasiswa Laki-laki. Untuk variabel Asal Daerah, terlihat didapatkan distribusi data yang cukup merata, yakni 44,7% mahasiswa berasal dari daerah Jawa Timur dan 55,3% berasal dari daerah luar Jawa Timur. Adapun pada variabel Status Kerja memperlihatkan bahwa mahasiswa Magister Statistika ITS sudah banyak yang memiliki status bekerja dengan prosentase 39% dan mahasiswa yang belum bekerja sebanyak 61%. Variabel Jalur Masuk yang ada pada data mahasiswa didominasi oleh jalur Mandiri dan jalur Instansi/Kerjasama, untuk jalur jenis beasiswa yang lain juga cukup merata penyebarannya. Untuk variabel Akreditasi asal S1 mahasiswa terlihat banyak yang berasal dari Perguruan Tinggi berakreditasi A sebesar 78%, dan prosentase sisanya berasal dari Perguruan Tinggi berakreditasi dengan predikat B dan C serta belum terakreditasi. Adapun pada variabel Linieritas Jurusan terlihat sangat didominasi oleh mahasiswa yang memiliki latar belakang Statistika dengan prosentase 63%, dan sisanya 31% berasal dari Matematika, 5% berlatarbelakang Pendidikan Matematika hanya 1% yang berasal dari jurusan lainnya. Penyebaran data pada setiap variabel masukan (*input*) memperlihatkan bahwa data penelitian memiliki tingkat variasi atau keragaman yang tinggi untuk diolah menjadi sebuah prediksi klasifikasi.

Pada variabel keluaran (*output*) terdapat 8 (delapan) kategori level performansi seperti yang sudah dijelaskan pada bab sebelumnya. Kedelapan kategori tersebut juga terdistribusi pada data penelitian yang disajikan Tabel 4.2 dan Gambar 4.2.

**Tabel 4.2** Distribusi Data Variabel Keluaran

<b>VARIABEL KELUARAN</b>	<b>JUMLAH</b>	<b>PROSENTASE</b>	
LEVEL KINERJA MAHASISWA	Level 0	19	6.0%
	Level 1	10	3.1%
	Level 2	22	6.9%
	Level 3	5	1.6%
	Level 4	76	23.9%
	Level 5	91	28.6%
	Level 6	41	12.9%
	Level 7	54	17.0%
<b>Total</b>	<b>318</b>	<b>100.0%</b>	



**Gambar 4.2** Grafik Distribusi Data Variabel Keluaran

Pada Tabel 4.2 dan Gambar 4.2 terlihat bahwa data hasil kinerja performansi mahasiswa Magister Statistika ITS yang dijadikan bahan penelitian didominasi oleh mahasiswa Level 5 dengan prosentase 29%. Dari 318 data yang berhasil dikumpulkan, 24% diantaranya masuk pada kategori Level 4, disusul Level 7 yakni level kinerja mahasiswa paling baik, sebanyak 17%. Adapun distribusi paling rendah adalah data mahasiswa yang berada di Level 3 dengan prosentasi hanya 1% saja. Level 3 merupakan level kinerja mahasiswa yang lulus dengan waktu tempuh lebih dari 4 (empat) semester namun memiliki IPK *cumlaude*. Kondisi mahasiswa pada level tersebut memang jarang ditemukan pada kasus yang ada di Program Studi Magister Statistika ITS.

#### **4.2. Implementasi *Support Vector Machine* pada Matlab**

Pada bagian ini akan dipaparkan mengenai implementasi *Support Vector Machine* menggunakan *Software* Matlab R2011b untuk mendapatkan prediksi kinerja mahasiswa sesuai dengan tujuan penelitian. Penggunaan Matlab R2011b dirasa sangat tepat digunakan karena *software* ini dirancang khusus untuk melakukan perhitungan matematis terkomputasi. Penerapan metode *Support Vector Machine* didasari dengan perhitungan matematis yang rumit dapat diimplementasikan dalam program dengan lebih mudah.

Matlab sudah memiliki *library* program perhitungan yang lengkap, termasuk juga untuk metode *Support Vector Machine*. Dalam penelitian ini, terdapat tiga *function* utama yang digunakan untuk membangun pemodelan prediksi kinerja mahasiswa. Tiga

*function* tersebut sudah tersedia dalam *library* Matlab, antara lain `svmtrain()`, `svmclassify()` dan `crossvalind()`.

Fungsi `svmtrain()` digunakan untuk melakukan proses *training* data menggunakan metode *Support Vector Machine* hingga didapatkan fungsi *hyperplane* sebagai pemisah antar kelas. Adapun tag `svmtrain()` *function* pada penelitian ini secara general ditampilkan dengan :

```
svmtrain(data_training, data_respon, 'kernel_function', fungsi_kernel);
```

Terdapat 4 (empat) *input* argumen yang perlu ditampilkan dalam fungsi. Argumen pertama `data_training` merupakan *array* data variabel masukan (*input*) yang digunakan dalam penelitian. *Array* ini memiliki 12 kolom sesuai dengan jumlah variabel masukan (*input*) yang sudah dijelaskan sebelumnya. Argumen kedua `data_training` adalah *array* data dengan 1 (satu) kolom yang berisi data variabel keluaran (*output*) level kinerja mahasiswa. Adapun argumen ketiga dan keempat merupakan argumen tambahan untuk fungsi Kernel sebagai fungsi matematis yang digunakan dalam metode *Support Vector Machine*. Argumen fungsi Kernel yang digunakan antara lain *Linear Kernel*, *Radial Basic Function (RBF) Kernel*, *Polynomial Kernel*, dan *Multilayer Perceptron (MLP) Kernel*.

Implementasi penggunaan masing-masing fungsi Kernel juga disesuaikan dengan parameter pada setiap fungsi Kernel yang dibutuhkan. Daftar fungsi Kernel yang diimplementasikan dalam Matlab dengan berbagai macam optimal parameternya ditampilkan pada Tabel 4.3.

**Tabel 4.3** Nilai Optimal Parameter pada Fungsi Kernel

Fungsi Kernel $K(x,y)$	Nilai Parameter Optimal	Implementasi Matlab
<b>Linear</b> $x^T y + c$	$c = 5$	<code>svmtrain(training, respon, 'kernel_function', 'linear', 'boxconstraint', 5);</code>
<b>Polynomial</b> $(x^T + c)^d$	$c = 5,$ $d = 1$	<code>svmtrain(training, respon, 'kernel_function', 'polynomial', 'boxconstraint', 5, 'polyorder', 1);</code>

Fungsi Kernel $K(x,y)$	Nilai Parameter Optimal	Implementasi Matlab
<b>Radial Basis Function (RBF)</b> $exp(-\gamma\ x - y\ ^2)$	$\gamma = 3.75$	<code>svmtrain(training, respon, 'kernel_function', 'rbf', 'rbf_sigma', 3.75);</code>
<b>Multilayer Perception (MLP)</b> $tan(p_1 x y' + p_2)$ dimana, $p_1 > 0$ dan $p_2 < 0$	$[p_1 \ p_2]$ = [1 -1]	<code>svmtrain(training, respon, 'kernel_function', 'mlp', 'mlp_params', [1 -1]);</code>

Fungsi `svmclassify()` merupakan fungsi yang digunakan untuk melihat prediksi data *testing* hasil dari proses implementasi *Support Vector Machine*. Hasil dari fungsi ini adalah klasifikasi yang sudah disesuaikan dengan proses `svmtrain()` sebelumnya. Berikut tag `svmclassify()` *function* pada penelitian ini secara general adalah :

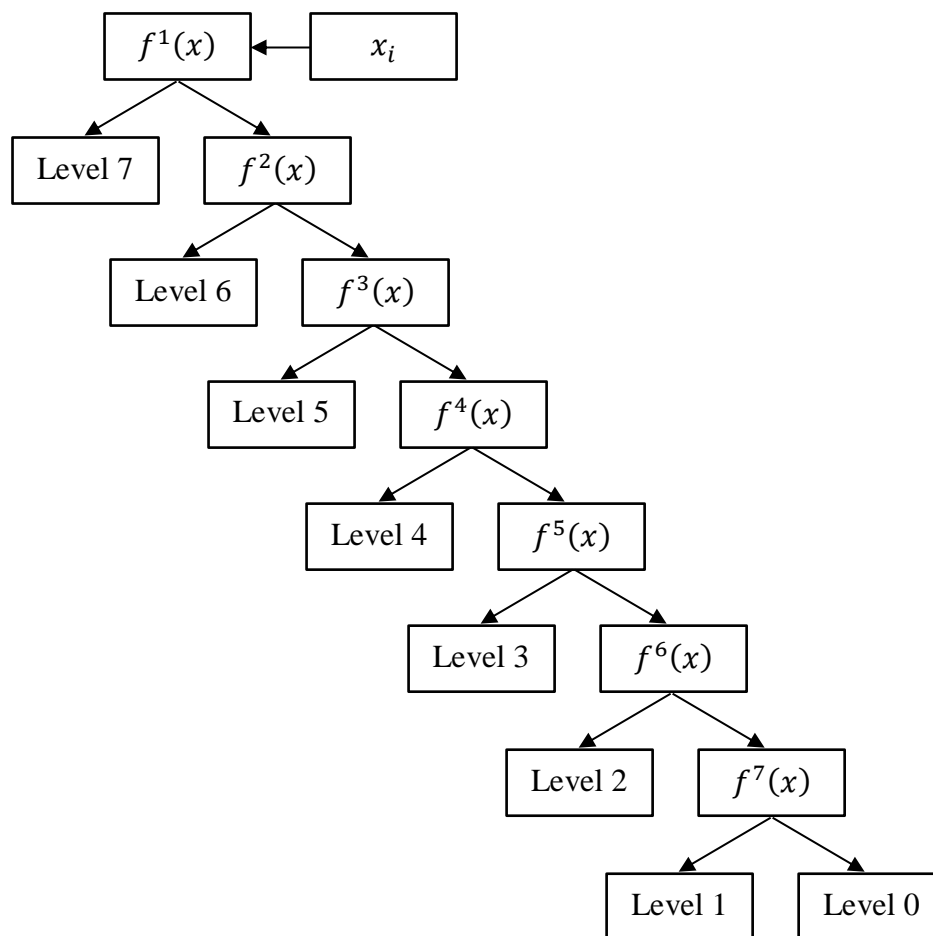
```
svmclassify(hasil_training, data_testing);
```

Terdapat 2 (dua) *input* argumen yang perlu ditampilkan dalam fungsi diatas. Argumen pertama `hasil_training` adalah data hasil proses training SVM yang sudah dilakukan sebelumnya menggunakan fungsi `svmtrain()`. Argumen kedua adalah `data_testing` yang merupakan data uji coba dalam penelitian. Argumen `data_testing` ini juga berbentuk *array* yang memiliki dimensi harus sama dengan data *training* sebelumnya, yakni 12 kolom. Hasil *output* dari fungsi `svmclassify()` ini merupakan prediksi kinerja mahasiswa yang dimaksud dalam penelitian.

Pada dasarnya metode *Support Vector Machine* ini hanya bisa digunakan untuk klasifikasi dengan dua kelas. Namun, dalam perkembangannya kemudian metode ini dapat digunakan untuk prediksi klasifikasi banyak kelas namun dengan penerapan algoritma yang sedikit berbeda. Adapun dalam penelitian ini, penggunaan *Support Vector Machine* untuk banyak kelas diimplementasikan dengan konsep *One Against All*, yakni Setiap model klasifikasi ke-*i* dilatih dengan menggunakan keseluruhan data, untuk mencari solusi. Adapun konsep *One Against All* SVM dengan banyak kelas khusus dalam penelitian ini ditampilkan dalam Tabel 4.4 dan Gambar 4.3.

**Tabel 4.4** Konsep *One Against All* SVM untuk Klasifikasi Level Mahasiswa

$y_i = 1$	$y_i = 0$	Hipotesis
Level 7	Bukan Level 7	$f^1(x) = (w^1)x + b^1$
Level 6	Bukan Level 6	$f^2(x) = (w^2)x + b^2$
Level 5	Bukan Level 5	$f^3(x) = (w^3)x + b^3$
Level 4	Bukan Level 4	$f^4(x) = (w^4)x + b^4$
Level 3	Bukan Level 3	$f^5(x) = (w^5)x + b^5$
Level 2	Bukan Level 2	$f^6(x) = (w^6)x + b^6$
Level 1	Bukan Level 1	$f^7(x) = (w^7)x + b^7$



**Gambar 4.3** Konsep *One Against All* SVM untuk Klasifikasi Level Mahasiswa

Setelah hasil prediksi kinerja mahasiswa didapatkan, selanjutnya dilakukan analisis hasil prediksi menggunakan teknik uji coba *k-fold Cross Validation* yang sebelumnya sudah dijelaskan pada bagian Rancangan Uji Coba.

Teknik uji coba ini sudah ada penggunaannya fungsinya dalam Matlab dengan *tag* fungsi `crossvalind()` sebagai berikut :

```
crossvalind('kfold', data_training, 10);
```

Fungsi `crossvalind()` sendiri merupakan fungsi untuk memberikan indeks *random* terhadap data yang digunakan dalam penelitian. Pemberian indeks disesuaikan dengan jumlah *k-fold* yang diinginkan. Terdapat 3 (tiga) argumen yang diperlukan untuk membangun fungsi `crossvalind()` antara lain `'kfold'`, `data_training`, dan jumlah *k* yang diinginkan. Argumen `'kfold'` digunakan untuk set teknik uji coba *k-fold cross validation*. Argumen kedua yakni `data_training` merupakan *array* data yang akan dilakukan pengujian menggunakan teknik uji coba tersebut. Adapun argumen ketiga adalah *value* atau nilai *k* yang merupakan banyaknya pembagian data, dimana dalam penelitian ini ditetapkan *k* bernilai 10. Nilai 10 didasarkan pada jumlah data sebanyak 318, dan *k*=10 merupakan nilai yang dirasa tepat untuk teknik uji coba tersebut.

Tiga fungsi *library* pada Matlab yang sudah dijabarkan diatas kemudian diimplementasikan dalam satu program beserta data yang sudah disiapkan sebelumnya, untuk pemodelan prediksi kinerja mahasiswa. Setelah program berhasil untuk dibangun sesuai dengan perencanaan teori sebelumnya, dilakukan pelatihan dan pengujian model tersebut.

#### **4.3. Perhitungan Ketepatan Prediksi**

Pada bagian ini akan dijelaskan mengenai konsep perhitungan ketepatan prediksi kinerja mahasiswa yang dibangun dengan implementasi metode *Support Vector Machine* (SVM). Perhitungan ketepatan prediksi dilakukan dengan menggunakan *confusion* matriks sebagai salah satu alat perhitungannya. *Confusion* matriks sendiri umumnya digunakan untuk melakukan perhitungan akurasi pada konsep data *mining* dengan membandingkan data hasil prediksi yang telah diproses dengan data aktual yang sudah ada. Metode ini melakukan perhitungan dengan 3 (tiga) keluaran, yaitu akurasi, presisi dan *recall*. Dalam penelitian ini *confusion* matriks yang dibangun sesuai dengan model dalam penelitian disajikan pada Tabel 4.5.



**Tabel 4.5** Konsep *Confusion* Matriks untuk *Output* Multikelas

		Hasil Prediksi				
		Level 0	Level 1	Level 2	...	Level 7
Data Aktual	Level 0	TP	FN	FN	FN	FN
	Level 1	FP	TP	FN	FN	FN
	Level 2	FP	FP	TP	FN	FN
	...	FP	FP	FP	TP	FN
	Level 7	FP	FP	FP	FP	TP

Tabel 4.5 menjelaskan bahwa *confusion* matriks yang dibangun khusus untuk penelitian dengan *output* banyak kelas. Seperti yang sudah diketahui sebelumnya bahwa pada penelitian ini, *output* prediksi kinerja mahasiswa dibagi menjadi 8 (delapan) level kinerja yakni dari level 0 hingga level 7. Data baris pada matriks ini diisi dengan data aktual sesuai dengan level kinerja yang sudah dimiliki oleh mahasiswa. Adapun kolom pada matriks dibangun dengan hasil prediksi level kinerja mahasiswa yang dikeluarkan oleh sistem SVM. Sinkronisasi baris dan kolom ini nantinya yang digunakan untuk mengisi isian matriks, sehingga dengan menggunakan *confusion* matriks yang dibangun dapat mempermudah analisis hasil prediksi kinerja mahasiswa sesuai dengan tujuan penelitian.

Adapun hasil prediksi level kinerja mahasiswa yang sesuai dengan data aktual akan diberi label TP (*True Positive*), yang artinya hasil uji coba prediksi level kinerja mahasiswa benar, tepat sesuai dengan aktualnya. Label FP (*False Positive*) didapatkan dengan menghitung hasil prediksi yang tidak sesuai dengan data aktual, namun level kinerja yang diprediksikan masih berada di bawah level kinerja yang sebenarnya. Label FP (*False Positive*) tersebut dipertimbangkan karena penelitian bertujuan untuk mendapatkan hasil prediksi kinerja mahasiswa yang nantinya digunakan sebagai sistem peringatan deteksi dini bagi pengelola program studi di perguruan tinggi. Sehingga, untuk pendeteksian level kinerja dengan hasil dibawah level aktual (data sebenarnya) dianggap sebagai kesalahan prediksi data yang bersifat positif (*False Positive*). Sebaliknya, hasil prediksi level kinerja mahasiswa yang berada di atas data level kinerja mahasiswa yang sebenarnya akan dianggap sebagai kesalahan prediksi data dengan sifat negatif, disebut sebagai *False Negative* (FN). Hasil TP, FP dan FN selanjutnya diolah menjadi prosentase Akurasi, Presisi dan *Recall* sesuai dengan persamaan rumus yang sudah dijelaskan pada Bab 2.

#### 4.4. Pelatihan dan Pengujian Model

Pelatihan dan pengujian model dibagi menjadi beberapa skenario utama. Masing-masing skenario memiliki tujuan yang berbeda dalam mengukur kinerja metode yang digunakan. Pengujian model klasifikasi menggunakan teknik *10-fold cross validation*, artinya 318 data dibagi menjadi 10 *fold* (bagian). Kombinasi 9 *fold* yang berbeda digabung dan digunakan sebagai data *training*, 1 *fold* sisa digunakan sebagai data *testing*. Pada bagian ini dipaparkan hasil uji coba setiap skenario untuk melihat performansi metode *Support Vector Machine* dalam Prediksi Kinerja Mahasiswa di Program Studi Magister Statistika ITS. Performansi prediksi diukur berdasarkan hasil perhitungan akurasi, presisi dan *recall* melalui *confusion* matriks yang sudah dijelaskan sebelumnya.

##### 4.4.1. Uji Performa Fungsi Kernel Terbaik

Pelatihan dan pengujian pada tahap ini memiliki tujuan untuk mendapatkan fungsi Kernel yang paling optimal guna diterapkan dalam pemodelan prediksi kinerja mahasiswa dalam penelitian ini. Fungsi Kernel yang akan digunakan dalam skenario uji coba diantaranya adalah *Linear Kernel*, *Polynomial Kernel*, *Radial Basis Function* (RBF) Kernel dan *Multilayer Perception* (MLP) Kernel. Hasil uji coba perhitungan ketepatan klasifikasi dilihat berdasarkan perhitungan akurasi, presisi dan *recall* pada setiap fungsi Kernel. Tabel 4.6 adalah hasil prosentase perhitungan akurasi dari uji coba implementasi SVM dengan berbagai fungsi Kernel-nya.

**Tabel 4.6** Hasil Prosentase Akurasi Uji Performa Fungsi Kernel

Fold ke-	Akurasi Fungsi Kernel			
	Linear	Polynomial	RBF	MLP
1	37.50%	41.93%	41.93%	18.75%
2	43.75%	43.75%	35.48%	28.12%
3	43.75%	43.75%	46.87%	25.00%
4	46.87%	46.87%	53.12%	9.37%
5	25.00%	31.25%	37.50%	19.35%
6	59.37%	37.50%	37.50%	15.62%
7	31.25%	31.25%	50.00%	6.25%
8	48.38%	37.50%	31.25%	28.12%
9	29.03%	48.38%	34.37%	12.50%
10	56.25%	50.00%	50.00%	12.90%
<b>Rata-rata</b>	<b>42.11%</b>	<b>41.21%</b>	<b>41.80%</b>	<b>17.60%</b>

Tabel 4.6 menunjukkan bahwa perhitungan akurasi prediksi kinerja mahasiswa menggunakan SVM dengan berbagai fungsi kernel memiliki nilai akurasi yang cukup baik. Dari perhitungan *10-fold cross validation* untuk akurasi tersebut, didapatkan rata-rata prosentase akurasi paling tinggi adalah prediksi SVM dengan fungsi *Linear Kernel* sebesar 42,11%. Prosentase akurasi paling baik kedua adalah implementasi SVM menggunakan *Radial Basis Function (RBF) Kernel* sebesar 41,80%, disusul penggunaan *Polynomial Kernel* dengan prosentase akurasi 41,21% dan *Multilayer Perception (MLP) Kernel* dengan rerata prosentase akurasi hanya 17,60%.

Selain perhitungan akurasi, uji coba juga dilakukan untuk melihat presisi dari proses prediksi kinerja mahasiswa menggunakan SVM dengan berbagai fungsi Kernel-nya. Tabel 4.7 menyajikan hasil prosentase perhitungan presisi uji coba prediksi kinerja mahasiswa dengan macam fungsi Kernel SVM.

**Tabel 4.7** Hasil Prosentase Presisi Uji Performa Fungsi Kernel

Fold ke-	Presisi Fungsi Kernel			
	Linear	Polynomial	RBF	MLP
1	80.00%	86.66%	76.47%	50.00%
2	87.50%	66.66%	57.89%	64.28%
3	77.77%	82.35%	83.33%	57.14%
4	88.23%	83.33%	77.27%	37.50%
5	61.53%	71.42%	57.14%	66.66%
6	82.60%	100.00%	57.14%	55.55%
7	83.33%	58.82%	88.88%	40.00%
8	88.23%	80.00%	66.66%	50.00%
9	69.23%	83.33%	57.89%	57.14%
10	81.81%	88.88%	80.00%	28.57%
<b>Rata-rata</b>	<b>80.02%</b>	<b>80.14%</b>	<b>70.27%</b>	<b>50.68%</b>

Hasil perhitungan rata-rata prosentase presisi dari uji coba presiksi kinerja mahasiswa yang ditampilkan pada Tabel 4.7 menunjukkan bahwa metode SVM cukup berhasil digunakan untuk prediksi tersebut. Terlihat nilai rata-rata presisi yang cukup tinggi dari hasil uji coba yang dilakukan menggunakan teknik *10-fold cross validation*. Hasil uji coba menunjukkan Fungsi *Polynomial Kernel* memiliki nilai rata-rata presisi yang paling baik dibandingkan dengan tiga fungsi Kernel yang lainnya. *Polynomial Kernel* berhasil memberikan kepresisian prediksi kinerja mahasiswa dengan SVM

sebesar 80,14%. Adapun fungsi Kernel yang lain memiliki hasil prosentase presisi dengan selisih yang tipis, yakni 80,02% untuk fungsi *Linear* Kernel. Adapun *Radial Basis Function* (RBF) Kernel berhasil memperoleh prosentase kepresisian sebesar 70,27% dan 50,68% untuk fungsi Kernel *Multilayer Perception* (MLP).

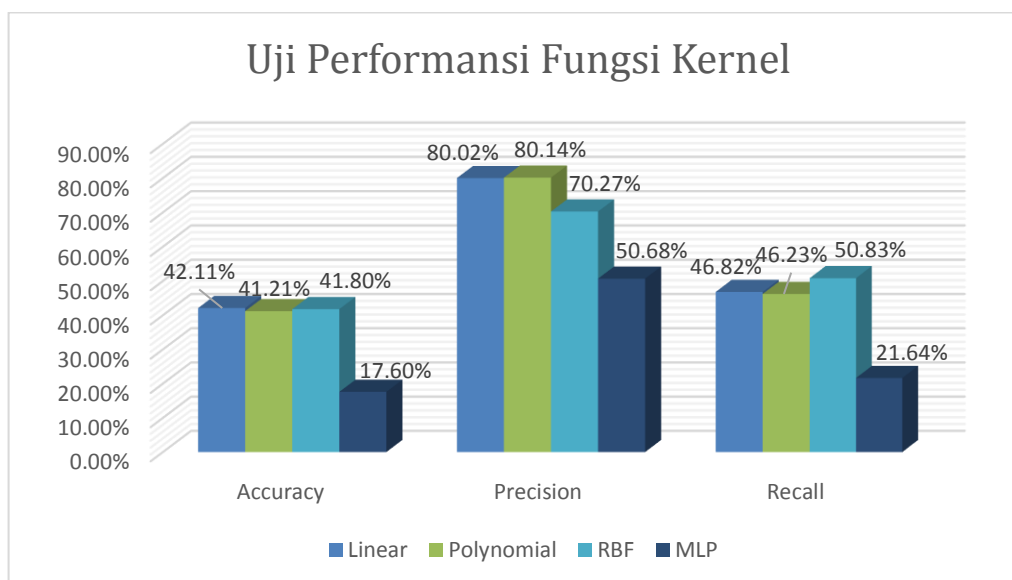
Selanjutnya pada Tabel 4.8 menyajikan hasil uji coba perhitungan rata-rata prosentase *recall* pada setiap implementasi fungsi Kernel SVM untuk prediksi kinerja mahasiswa Program Studi Magister Statistika ITS. *Recall* adalah salah satu pengukuran keberhasilan prediksi untuk melihat tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

**Tabel 4.8** Hasil Prosentase *Recall* Uji Performa Fungsi Kernel

Fold ke-	Recall Fungsi Kernel			
	Linear	Polynomial	RBF	MLP
1	41.37%	44.82%	48.14%	23.07%
2	46.66%	56.00%	47.82%	33.33%
3	50.00%	48.27%	51.72%	30.76%
4	50.00%	51.72%	62.96%	11.11%
5	29.62%	35.71%	52.17%	21.42%
6	67.85%	37.50%	52.17%	17.85%
7	33.33%	40.00%	53.33%	6.89%
8	51.72%	41.37%	37.03%	39.13%
9	33.33%	53.57%	45.83%	13.79%
10	64.28%	53.33%	57.14%	19.04%
<b>Rata-rata</b>	<b>46.82%</b>	<b>46.23%</b>	<b>50.83%</b>	<b>21.64%</b>

Dari Tabel 4.8 terlihat bahwa hasil prosentase perhitungan *recall* menghasilkan nilai yang tidak terlalu tinggi jika dibandingkan dengan nilai akurasi maupun presisi. Sebuah sistem prediksi akan dianggap baik jika tingkat *recall* maupun presisinya tinggi. Pada pelatihan dan pengujian yang sudah dilakukan, nilai prosentase *recall* yang paling baik didapatkan oleh implementasi SVM dengan penggunaan *Radial Basis Function* (RBF) Kernel sebesar 50,83%. Hasil Fungsi Kernel lainnya tampak memiliki prosentase *recall* yang tidak lebih dari 50%. *Linear* Kernel memiliki *record* hasil perhitungan *recall* dengan rata-rata 46,82%, *Polynomial* Kernel berada di posisi ketiga dengan prosentase 46,23% dan hasil terendah didapatkan oleh *Multilayer Perception* Kernel dengan prosentase *recall* sebesar 21,64%. Adapun didapatkan rekapitulasi perhitungan akurasi,

presisi dan *recall* dari uji coba performa fungsi Kernel untuk Prediksi Kinerja Mahasiswa menggunakan SVM dalam grafik Gambar 4.4.



**Gambar 4.4** Grafik Hasil Uji Coba Performansi Fungsi Kernel

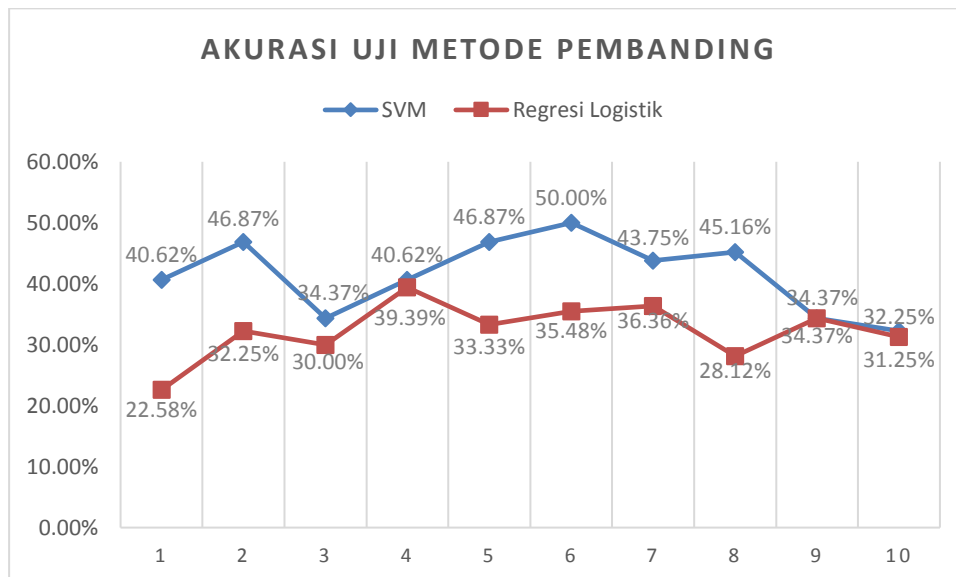
Grafik pada Gambar 4.4 menjelaskan bahwa untuk hasil uji coba performansi Fungsi Kernel yang diimplementasikan pada prediksi kinerja mahasiswa menggunakan *Support Vector Machine* terlihat paling optimal adalah *Linear* Kernel. Jika dilihat dari hasil akurasi, presisi dan *recall* memang performa *Linear* Kernel tidak jauh berbeda dengan fungsi Kernel lainnya, namun penggunaan *Linear* Kernel memiliki performa yang lebih stabil dalam mendapatkan perhitungan akurasi, presisi dan *recall*. Hasil yang didapatkan nampak berbeda jika dibandingkan dengan penelitian Kaur, P. dkk (2005) yang mengatakan bahwa fungsi Kernel paling optimal untuk prediksi kinerja adalah *Multilayer Perception* Kernel.

#### 4.4.2. Uji Performa dengan Metode Pemanding

Pada bagian ini akan dipaparkan mengenai hasil uji coba untuk melihat performa ketepatan klasifikasi SVM dibandingkan dengan metode pemanding Regresi Logistik. Metode Regresi Logistik yang merupakan metode statistik konvensional dirasa tepat digunakan sebagai metode pemanding, karena metode tersebut sudah banyak dilakukan dalam berbagai penerapan di segala bidang. *Support Vector Machine* sendiri merupakan metode pengklasifikasian yang diklaim dapat menjadi metode yang memiliki hasil yang

global optimal dibandingkan dengan metode statistik konvensional biasa layaknya Regresi Logistik.

Regresi Logistik sebagai metode pembandingan dalam penelitian ini juga diterapkan menggunakan software Matlab R2011b. Adapun penerapannya juga menggunakan fungsi *library* yang sudah ada pada Matlab. Uji coba juga dilakukan dengan teknik *10-fold Cross Validation* yang kemudian dibandingkan hasil prediksinya dengan metode *Support Vector Machine (SVM)*. Sama seperti pelatihan dan pengujian sebelumnya, uji coba performa dengan metode pembandingan Regresi Logistik ini juga dilakukan dengan perhitungan hasil akurasi, presisi dan *recall*. Grafik prosentase hasil perhitungan akurasi prediksi kinerja mahasiswa menggunakan metode SVM dan Regresi Logistik yang ditampilkan pada Gambar 4.5.

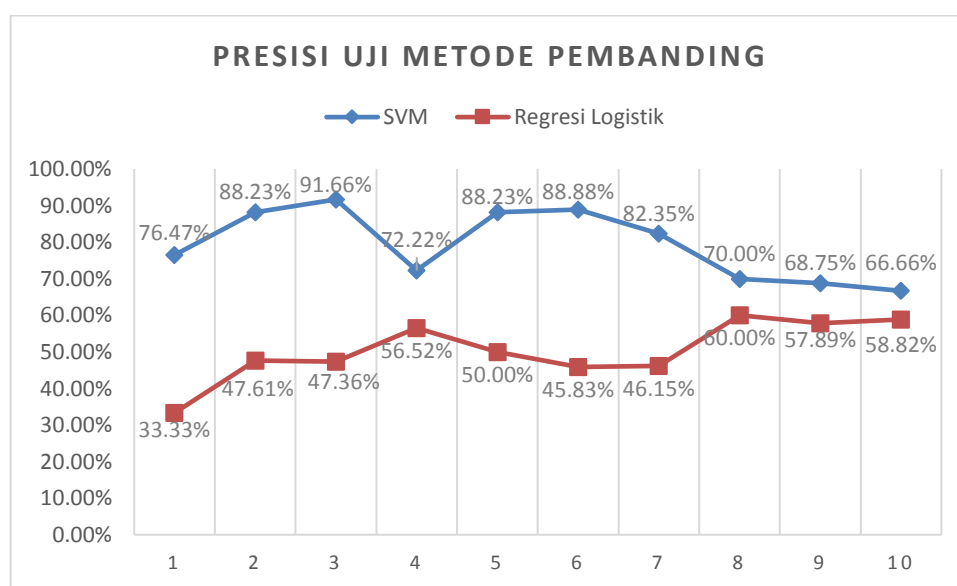


**Gambar 4.5** Grafik Hasil Akurasi dengan SVM dan Regresi Logistik

Gambar 4.5 memperlihatkan hasil perhitungan akurasi prediksi kinerja mahasiswa menggunakan SVM dan metode pembandingnya, yakni Regresi Logistik. Sumbu (x) pada grafik menerangkan *k-fold* yang dilakukan dalam percobaan sebanyak 10 (sepuluh). Sedangkan sumbu (y) pada grafik menerangkan prosentase akurasi yang didapatkan pada hasil uji coba setiap *k-fold*. Dari Gambar 4.5 dapat dilihat bahwa perhitungan akurasi yang dihasilkan dengan *Support Vector Machine (SVM)* memiliki nilai prosentase yang lebih tinggi dibandingkan dengan metode pembandingnya. Adapun hasil uji *10-fold cross validation* ini menghasilkan rerata sebesar 41,49% untuk

metode SVM dan 32.31% untuk metode Regresi Logistik. Dari uji akurasi untuk kedua metode tersebut dapat disimpulkan bahwa prediksi kinerja mahasiswa menggunakan metode *Support Vector Machine* (SVM) memiliki nilai akurasi yang lebih tinggi jika dibandingkan dengan metode Regresi Logistik.

Selanjutnya untuk melihat hasil kepresisian prediksi kinerja mahasiswa dengan dua metode, Gambar 4.6 menyajikan grafik hasil perhitungan presisi kedua metode yang dibandingkan.

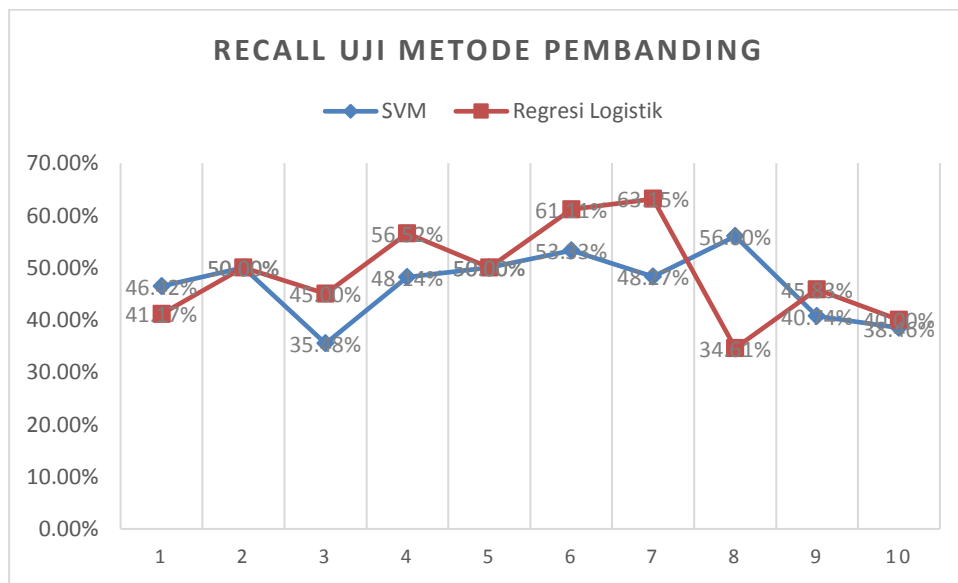


**Gambar 4.6** Grafik Hasil Presisi dengan SVM dan Regresi Logistik

Pada grafik yang disajikan Gambar 4.6 memperlihatkan bahwa perhitungan prosentase kepresisian uji coba untuk penggunaan SVM jauh melebihi penggunaan Regresi Logistik. Prediksi kinerja mahasiswa menggunakan *Support Vector Machine* (SVM) menghasilkan rerata kepresisian sebesar 79,34%. Sedangkan metode pembandingan dengan statistik konvensional yakni regresi logistik menghasilkan rerata presisi sebesar 50,35%. Namun, jika dilihat dari grafik pada setiap *k-fold* diketahui bahwa metode Regresi Logistik memiliki garis kestabilan yang lebih baik dan cenderung meningkat jika dibandingkan dengan hasil penggunaan metode *Support Vector Machine* (SVM). Hasil presisi pada SVM memiliki nilai yang cenderung tidak stabil, perbedaan perhitungan presisi yang dihasilkan pada setiap *k-fold* cukup signifikan. Kestabilan performansi juga perlu dijadikan tolak ukur dalam penelitian, karena prediksi kinerja mahasiswa memerlukan performa kestabilan yang baik dalam menghasilkan klasifikasi

atau prediksi yang diinginkan. Namun bagaimanapun, tidak dapat dipungkiri bahwa Prediksi Kinerja Mahasiswa menggunakan metode *Support Vector Machine* (SVM) masih lebih handal digunakan jika dibandingkan dengan metode Regresi Logistik karena nilai kepresisian yang tinggi.

Selain ukuran akurasi dan prediksi, ketepatan prediksi kinerja juga diukur melalui hasil perhitungan *recall* yang didapatkan dari implementasi kedua metode. Gambar 4.7 Grafik Uji *Recall* untuk prediksi kinerja mahasiswa menggunakan metode *Support Vector Machine* (SVM) dan Regresi Logistik.



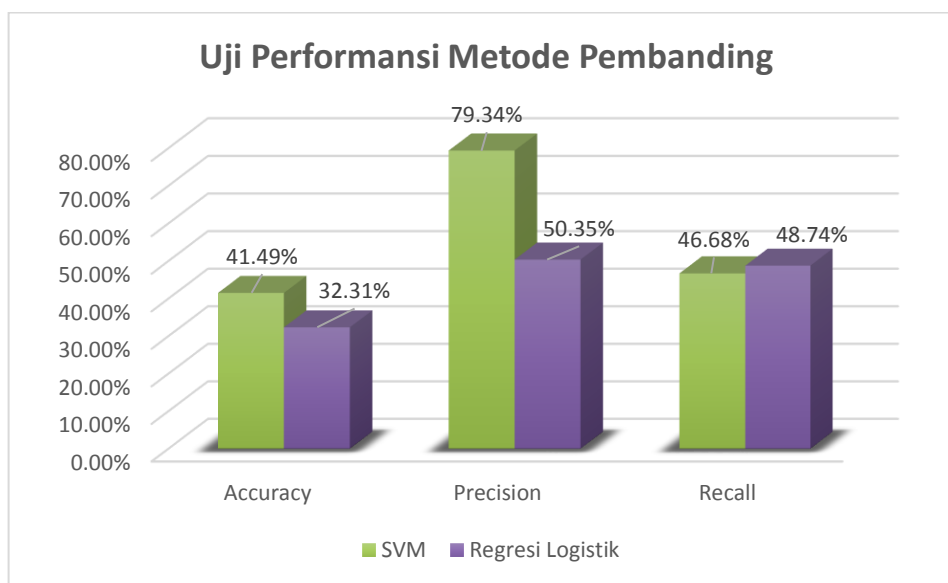
**Gambar 4.7** Grafik Hasil *Recall* Metode SVM dan Regresi Logistik

Grafik hasil perhitungan *recall* dari penelitian dengan dua metode pada Gambar 4.7 memperlihatkan bahwa penggunaan SVM dan Regresi Logistik pada prediksi kinerja mahasiswa menghasilkan rerata *recall* yang sama-sama tidak lebih dari 50%. Hal tersebut terlihat bahwa dari 10-fold percobaan yang dilakukan, penggunaan metode Regresi Logistik menghasilkan rerata *recall* yang lebih tinggi yakni sebesar 48,74%. Adapun penerapan prediksi kinerja mahasiswa dengan metode SVM menghasilkan *recall* yang sedikit lebih rendah, yakni sebesar 46,68%. Perbedaan prosentase rerata perhitungan *recall* yang tidak signifikan mengindikasikan penerapan SVM masih lebih baik. Adapun jika dilihat dari grafik pada setiap *fold* dapat dilihat bahwa metode SVM memiliki hasil perhitungan *recall* yang lebih stabil dibandingkan dengan metode Regresi Logistik. Pada hasil beberapa *fold* memang metode Regresi Logistik nampak lebih



unggul perhitungan *recall* yang dihasilkan, namun pada *fold* ke-8 terlihat bahwa perhitungan *recall* regresi logistik turun drastis ke angka 34,61%. Sehingga dapat dikatakan bahwa hasil prediksi menggunakan regresi logistik masih memiliki tingkat stabilitas performa yang rendah pada perhitungan *recall*-nya.

Dari hasil perhitungan akurasi, presisi dan *recall* yang didapatkan dari pengujian penerapan kedua metode dengan menggunakan teknik *10-fold cross validation*, didapatkan rerata akurasi, presisi dan *recall* yang ditampilkan pada grafik Gambar 4.8.



**Gambar 4.8** Grafik Hasil Uji Coba Performansi Metode Pemanding

Grafik pada Gambar 4.8 menampilkan hasil uji coba perbandingan performansi penggunaan SVM dan Regresi Logistik pada prediksi kinerja mahasiswa. Adapun dari hasil tersebut, performa akurasi, presisi maupun *recall* tampak metode *Support Vector Machine* (SVM) menghasilkan performa yang lebih baik. Rata-rata nilai akurasi dan presisi dari penggunaan metode *Support Vector Machine* (SVM) masing-masing bernilai 41,49% dan 79,34%. Prosentase nilai tersebut jauh melebihi nilai yang dihasilkan dari penerapan metode pembandingnya, yakni Regresi Logistik. Untuk perhitungan *recall* dari Regresi Logistik memang nampak lebih tinggi, namun perbedaan yang dihasilkan tidak terlalu signifikan dari nilai *recall* yang dimiliki oleh metode SVM. Rata-rata *recall* yang dihasilkan dari uji coba untuk SVM sebesar 46,68% dan untuk Regresi Logistik sedikit lebih tinggi yakni sebesar 48,74%. Sehingga dapat disimpulkan bahwa dari hasil uji coba yang dilakukan dalam kasus penelitian ini, Prediksi Kinerja Mahasiswa

menggunakan metode *Support Vector Machine* (SVM) memiliki performa yang lebih baik dibandingkan dengan prediksi kinerja yang menggunakan metode Regresi Logistik.

#### **4.5. Analisis Hasil Penerapan SVM**

Pada bagian ini, penulis akan menyajikan analisis hasil penerapan metode *Support Vector Machine* (SVM) pada prediksi kinerja mahasiswa Magister Statistika ITS. Prediksi ini diharapkan dapat membantu pengelola program studi untuk mendeteksi secara dini performansi kinerja mahasiswanya ketika berada pada awal perkuliahan. Sebuah prediksi yang baik tentunya harus memiliki keakuratan yang tinggi. Analisis hasil penerapan SVM dipaparkan menjadi beberapa bagian, yakni analisis berdasarkan ketepatan *False Positive* (FP) dan *False Negative* (FN), analisis berdasarkan pengelompokan level performansi mahasiswa, serta analisis pengaruh variabel masukan (*input*) terhadap kinerja performansi mahasiswa.

##### **4.5.1. Analisis *False Positive* (FP) dan *False Negative* (FN)**

Dalam penelitian prediksi kinerja mahasiswa dengan studi kasus Magister Statistika ITS ini, sebuah prediksi dikatakan bersifat positif apabila kinerja mahasiswa yang diprediksikan memiliki level kinerja yang lebih rendah dari pada data aktualnya. Sebagai contoh, ketika sistem melakukan prediksi kinerja seorang mahasiswa, sistem tersebut menghasilkan prediksi Level 2 sedangkan secara aktual mahasiswa tersebut berada pada level kinerja ke-5. Hasil prediksi yang didapatkan dari sistem termasuk prediksi yang bernilai positif. Adapun apabila prediksi menunjukkan Level 7 padahal secara aktual mahasiswa tersebut berada pada posisi level 5, artinya prediksi yang dilakukan oleh sistem bernilai negatif. Hasil prediksi kinerja mahasiswa yang dilakukan dengan teknik uji coba *10-fold Cross Validation* dengan membandingkan hasil prediksi dan data aktual yang sebenarnya, ditampilkan pada Tabel 4.9.

**Tabel 4.9** *Confusion* Matriks Hasil Prediksi Kinerja Mahasiswa

		Hasil Prediksi							
		Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Data Aktual	Level 0	1	1	1	-	2	5	6	3
	Level 1	-	-	2	-	2	5	1	-
	Level 2	-	1	4	-	12	-	2	3
	Level 3	-	-	-	1	-	1	2	1
	Level 4	-	-	4	1	25	15	17	14
	Level 5	-	-	3	-	6	38	11	33
	Level 6	-	-	1	-	3	2	20	15
	Level 7	-	-	-	-	3	3	4	44

Prediksi kinerja mahasiswa digunakan sebagai deteksi dini untuk pengelola program studi dalam melihat kinerja mahasiswanya. Maka dari itu, prediksi dengan hasil yang berada dibawah data aktual dikatakan lebih baik, pun sebaliknya, prediksi yang berada diatas data aktual dikatakan sebagai kesalahan yang bersifat negatif. Prediksi kinerja mahasiswa dalam kasus ini diharapkan dapat memiliki nilai *False Positive* (FP) yang lebih tinggi dibandingkan nilai *True Positive* (TP) dan *False Negative* (FN). Hal tersebut dipertimbangkan karena prediksi yang memiliki FN lebih banyak menghasilkan sistem warning dengan kesalahan fatal dalam kasus klasifikasi kinerja mahasiswa ini meskipun hasil TP juga sudah baik.

Terlihat dari Tabel 4.9, dari 318 data pelatihan, terdapat banyak prediksi yang meleset melebihi data aktual dari level kinerja mahasiswa. Dalam hasil uji coba, memang sudah cukup banyak prediksi yang dihasilkan akurat sesuai dengan data aktual, yakni sebanyak 133 mahasiswa diprediksi secara benar dan akurat. Namun, terdapat hasil prediksi dengan akurasi kesalahan bersifat negatif, yang menyatakan prediksi berlevel lebih tinggi dibandingkan data sebenarnya. Sebagai contoh, terdapat 33 mahasiswa yang diprediksi berpotensi memiliki level kinerja mahasiswa ke-7, ternyata secara aktual 33 mahasiswa tersebut berada pada level kinerja ke-5. Adapun secara keseluruhan, terdapat 154 data mahasiswa yang memiliki hasil kesalahan prediksi bersifat negatif (*false negative*), dan hanya ada 31 mahasiswa yang memiliki kesalahan prediksi bersifat positif (*false positive*). Dari analisis hasil tersebut, penggunaan metode *Support Vector Machine* (SVM) untuk prediksi kinerja mahasiswa perlu dilakukan penyempurnaan dengan optimasi pengaturan parameter untuk mendapatkan *false positive* yang lebih maksimal.

Adapun setelah uji coba optimasi parameter dilakukan, didapatkan bahwa SVM dengan fungsi *Radial Basis Fuction* (RBF) Kernel dan parameter sigma sebesar 1.75 dapat meningkatkan hasil *false positive* yang ada dalam penelitian. Tabel 4.10 menunjukkan hasil prediksi kinerja mahasiswa menggunakan SVM dengan optimasi parameter yang dimaksud.

**Tabel 4.10** *Confusion* Matriks Hasil Prediksi Kinerja Mahasiswa dengan Optimasi FP

		Hasil Prediksi							
		Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Data Aktual	Level 0	12	0	0	1	2	1	2	1
	Level 1	5	2	1	0	1	1	0	0
	Level 2	4	1	3	0	7	6	1	0
	Level 3	2	0	0	0	0	1	1	1
	Level 4	14	1	5	0	19	14	15	8
	Level 5	14	2	4	0	11	42	3	15
	Level 6	7	0	0	0	9	1	20	4
	Level 7	8	0	0	0	6	12	5	23

Tabel 4.10 memperlihatkan bahwa dengan optimasi parameter SVM yang sudah dilakukan, tingkat akurasi hasil prediksi memang menurun. Dari 318 data yang digunakan, hanya terdapat 119 data dengan prediksi akurasi yang tepat. Namun, hasil prediksi kesalahan bersifat negatif menjadi lebih sedikit dibandingkan dengan hasil prediksi kesalahan positif. Terdapat 111 data prediksi yang bersifat *false positive* dan 86 data prediksi yang bersifat *false negative*. Hasil prediksi tersebut dapat dikatakan lebih baik dibandingkan uji coba sebelumnya, karena prediksi yang dihasilkan dapat meminimalisir hasil prediksi yang melebihi data aktual dari kinerja mahasiswa sebenarnya.

#### 4.5.2. Analisis Pengelompokan Level Kinerja Mahasiswa

Selain perhitungan sesuai dengan level kinerja mahasiswa yang sudah ditentukan, hasil prediksi level kinerja juga dikelompokkan lagi berdasarkan performansi mahasiswa tersebut. Adapun pertimbangan pengelompokan performansi mahasiswa disesuaikan dengan kedekatan level kinerja mahasiswa yang sudah dimodelkan sebelumnya sebagai variabel keluaran (*output*). Tabel 4.11 menyajikan tentang pengelompokan performansi mahasiswa yang lebih luas sesuai dengan kedekatan level kerjanya.

**Tabel 4.11** Kelompok Performansi Mahasiswa

<b>Kelompok Performansi</b>	<b>Level Kinerja</b>
Kurang Baik	Level 0
	Level 1
	Level 2
	Level 3
Baik	Level 4
	Level 5
	Level 6
	Level 7

Mahasiswa dengan prediksi kinerja Level 0, 1, 2 dan Level 3 dikelompokkan menjadi mahasiswa dengan performansi yang Kurang Baik. Level 4, 5, 6 dan 7 merupakan level kinerja dengan kelompok performansi Baik. Pertimbangan yang diperhitungkan dalam pembagian kelompok performansi ini adalah Level 0 merupakan level performansi mahasiswa yang diberi status Dropout ataupun mengundurkan diri di tengah perkuliahan. Selain itu, Level 1, 2 dan 3 merupakan level performansi mahasiswa yang memerlukan masa studi lebih dari empat semester untuk menyelesaikan perkuliahannya. Ke-empat level tersebut merupakan level performansi mahasiswa yang kurang baik sehingga perlu dipertimbangkan oleh pengelola program studi dan dijadikan kelompok performansi tersendiri. Adapun kelompok performansi dengan label Baik merupakan kelompok mahasiswa yang bisa menyelesaikan studi tepat dan kurang dari empat semester. Hasil akurasi prediksi performansi mahasiswa ini juga digunakan sebagai salah satu tolak ukur keberhasilan sistem dalam melakukan prediksi kinerja mahasiswa pada penelitian.

Pada bagian ini akan dijelaskan mengenai analisis penerapan SVM berdasarkan akurasi pengelompokan level kinerja. Pada Tabel 4.11 sudah dijelaskan bahwa diperlukan pengelompokan level kinerja untuk mengetahui akurasi yang dihasilkan berdasarkan kedekatan levelnya. Prediksi yang sudah dilakukan sebelumnya, merupakan prediksi dengan tingkat level performansi dimana hasil akurasi dirasa masih kurang.

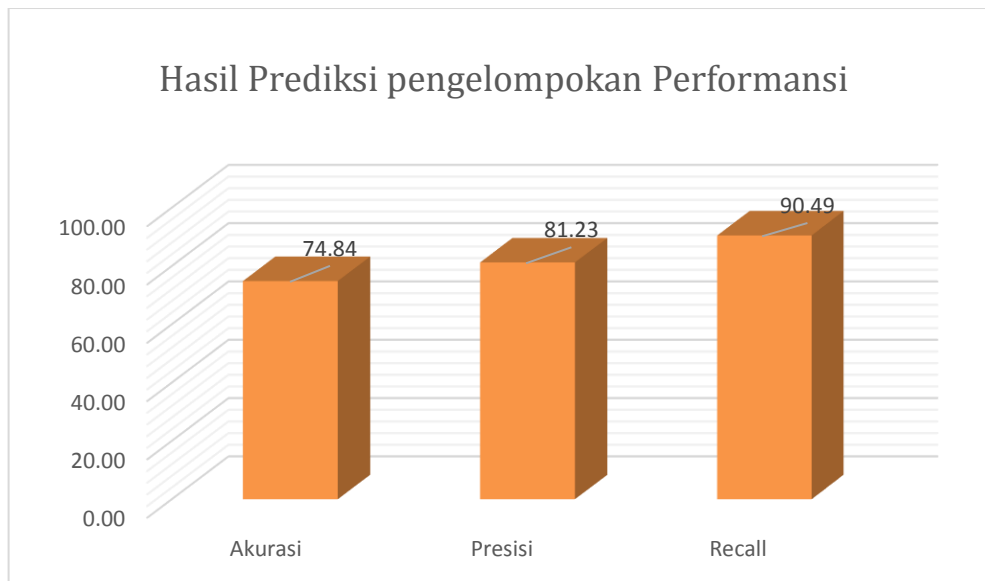
Penulis mencoba untuk mengelompokkan hasil prediksi dengan banyak pertimbangan. Salah satu pertimbangan yakni akurasi level kinerja yang sudah dilakukan sebelumnya banyak meleset ke prediksi level kinerja yang tidak terlalu signifikan dengan data aktual. Sebagai contoh, 14 data mahasiswa diprediksi memiliki kinerja

performansi mahasiswa di level 5, padahal secara aktual kelima belas mahasiswa tersebut berada pada performansi mahasiswa di level 4. Penulis berpendapat bahwa level 4 dan level 5 dalam prediksi kinerja mahasiswa ini memiliki perbedaan yang tidak terlalu signifikan, dalam artian dapat dikelompokkan menjadi satu performansi mahasiswa. Sehingga, disajikan hasil prediksi kinerja mahasiswa menggunakan SVM yang sudah dikelompokkan berdasarkan jenis performansinya pada Tabel 4.12. Adapun hasil prediksi sudah menggunakan optimasi SVM dengan fungsi RBF Kernel berparameter sigma 1,75.

**Tabel 4.12** Prediksi Kinerja Mahasiswa dengan Pengelompokan Level Performansi

		Hasil Prediksi								
		Kurang Baik				Baik				
		Lvl 0	Lvl 1	Lvl 2	Lvl 3	Lvl 4	Lvl 5	Lvl 6	Lvl 7	
<b>Data Aktual</b>	<b>Kurang Baik</b>	Level 0	12	-	-	1	2	1	2	1
		Level 1	5	2	1	-	1	1	-	-
		Level 2	4	1	3	-	7	6	1	-
		Level 3	2	-	-	-	-	1	1	1
	<b>Baik</b>	Level 4	14	1	5	-	19	14	15	8
		Level 5	14	2	4	-	11	42	3	15
		Level 6	7	-	-	-	9	1	20	4
		Level 7	8	-	-	-	6	12	5	23

Tabel 4.12 menyajikan hasil prediksi kinerja mahasiswa menggunakan SVM yang dikelompokkan menjadi 2 kelompok besar performansi, yakni Kurang Baik untuk level 0, 1, 2 dan 3, serta kelompok Baik untuk level 4 hingga level 7. Jika dilihat berdasarkan pengelompokan besar performansinya, dari 318 data mahasiswa yang diuji, terdapat 31 data yang akurat diprediksi berada pada kelompok Kurang Baik dan 207 data akurat diprediksi pada kelompok Baik. Sehingga total data yang diprediksi secara benar dan tepat sesuai dengan kelompok performansi adalah 238 data. Adapun 80 data sisanya berada pada posisi prediksi yang tidak akurat. Adapun hasil perhitungan akurasi, presisi dan *recall* setelah dilakukan pengelompokan performansi disajikan pada grafik Gambar 4.9.



**Gambar 4.9** Grafik Hasil Prediksi Pengelompokan Performansi Kinerja

Dengan analisis berdasarkan kelompok besar ini, ketepatan prediksi berdasarkan klasifikasi performansi menjadi jauh lebih baik, jika dibandingkan dengan klasifikasi untuk setiap level kinerja. Terlihat pada Gambar 4.9 hasil grafik prediksi performansi dengan pengelompokannya, prosentase akurasi sebesar 74,84%, presisi sebesar 81,23% dan prosentase *recall* sebesar 90,49%.

Sebagai pengelola program studi yang menjadi pengguna, sistem yang dibangun dengan metode *Support Vector Machine* (SVM) sudah cukup dapat mewakili prediksi kinerja mahasiswa guna mendeteksi secara dini kinerjanya di awal perkuliahan dan sebagai bahan pertimbangan pengambilan keputusan. Adapun jika dilihat lebih detail, banyak data mahasiswa yang diprediksi berada pada level 6 dan 7 namun secara aktual hanya berada pada level 4 ataupun 5. Hal ini mengindikasikan bahwa banyak mahasiswa yang sebenarnya berpotensi berada pada level performansi yang tinggi hingga di akhir masa perkuliahan, hanya saja karena satu dan lain hal maka mahasiswa tersebut cukup puas berada pada level performansi 4 dan 5. Hal tersebut dapat menjadi tugas pengelola program studi untuk lebih memaksimalkan kinerja performansi mahasiswanya hingga menghasilkan lulusan program studi yang berkualitas.

#### **4.5.3. Analisis Pengaruh Variabel Masukan untuk Prediksi Kinerja Mahasiswa**

Selanjutnya, penulis mencoba untuk menganalisis variabel masukan (*input*) yang berpengaruh signifikan terhadap kinerja mahasiswa sebagai variabel keluaran (*output*).

Hal ini bertujuan untuk membantu pengelola program studi meningkatkan kinerja performansi lulusan mahasiswanya, hingga mendapatkan performansi yang maksimal. Berikut penulis akan membahas satu per satu mengenai signifikasi variabel masukan terhadap level performansi kinerja mahasiswa.

Setelah dilakukan uji coba pengaruh perubahan variabel, diketahui bahwa variabel Jenis Kelamin dan Umur yang ada dalam data mahasiswa tidak berpengaruh signifikan terhadap hasil akhir kinerja performansi mahasiswa. Hal tersebut berbeda dengan penelitian yang dihasilkan oleh Asogwa, O. C (2015) yang mengatakan bahwa Jenis Kelamin dan Umur berpengaruh signifikan terhadap kinerja mahasiswa menggunakan algoritma ANN. Dalam penelitian ini, perbedaan Jenis Kelamin maupun Umur mahasiswa tidak berpengaruh signifikan terhadap kinerja mahasiswa yang ada di Program Studi Statistika ITS.

Adapun variabel Asal Daerah yang didapatkan dari data mahasiswa ternyata berpengaruh signifikan terhadap performansi kinerja mahasiswa. Dari percobaan pengaruh perubahan variabel terhadap prediksi kinerja mahasiswa menggunakan SVM, terlihat bahwa mahasiswa yang berasal dari daerah luar Jawa Timur memiliki performansi yang lebih baik ketimbang mahasiswa yang berasal dari daerah Jawa Timur. Untuk variabel Status Kerja memiliki signifikansi yang cukup besar. Jika data mahasiswa memiliki status sudah bekerja di usia berapapun, sistem memprediksi mahasiswa tersebut akan memiliki kinerja performansi yang baik hingga mahasiswa tersebut lulus dari perkuliahan. Namun, ketika data mahasiswa menunjukkan usia 23-24 dan memiliki status Belum Bekerja, maka sistem akan memprediksi bahwa mahasiswa tersebut berada pada level performansi yang kurang baik. Jika usia mahasiswa ketika masuk dalam perkuliahan Magister berada pada posisi diatas 24 tahun dengan status belum bekerja, performansi mahasiswa diprediksi berada pada kisaran level 4 hingga level 6. Hanya saja, akan lebih baik jika pengelola program studi lebih mempertimbangkan mahasiswa yang memiliki status Sudah Bekerja.

Variabel Jalur Masuk bagi mahasiswa yang melanjutkan ke jenjang Magister Program Studi Statistika juga berpengaruh signifikan terhadap kinerja performansinya. Mahasiswa yang masuk melalui jalur Mandiri dan *Joint Degree* akan memiliki kinerja performansi yang lebih baik jika dibandingkan dengan mahasiswa yang masuk melalui jalur beasiswa, *fresh graduate*, maupun jalur Kerjasama. Namun, penurunan level



performansi untuk variabel ini tidak begitu signifikan, sehingga masih bisa berada pada kategori performansi yang cukup baik.

Variabel nilai Toefl bagi mahasiswa juga cukup berpengaruh signifikan terhadap hasil performansi. Ditinjau dari peraturan Akademik ITS, salah satu syarat mahasiswa dinyatakan lulus dari jenjang Magister adalah memiliki nilai Toefl minimal sebesar 477. Jika dilakukan uji coba pada sistem, prediksi kinerja mahasiswa dapat menurun apabila pada awal masuk perkuliahan mahasiswa tersebut hanya memiliki nilai Toefl tidak lebih dari 388. Sistem memprediksi adanya level penurunan kinerja jika nilai toefl mahasiswa berada pada posisi kurang dari 388. Apabila pada awal perkuliahan mahasiswa sudah memiliki nilai toefl yang lebih dari itu, mahasiswa tersebut diprediksi aman dan baik level kinerjanya. Sama seperti Toefl, Variabel nilai TPA juga memiliki pengaruh yang signifikan terhadap kinerja performansi mahasiswa. TPA dengan nilai tidak lebih dari 391 juga mempengaruhi penurunan performansi mahasiswa di akhir perkuliahan. Sehingga pengelola program studi baiknya membatasi nilai minimum TPA berkisar 400 supaya dapat menghindari hasil kinerja mahasiswa yang rendah di akhir kelulusan nanti.

Adapun variabel Akreditasi Asal Perguruan Tinggi juga berpengaruh signifikan terhadap hasil prediksi performansi mahasiswa. Jika pengelola program studi menerima mahasiswa yang berasal dari perguruan tinggi yang terakreditasi selain A, penurunan performansi mahasiswa banyak didapatkan. Dari hasil data *training* yang sudah dilakukan uji coba pelatihan, asal perguruan tinggi yang terakreditasi B, C dan belum terakreditasi selalu menghasilkan kualitas mahasiswa dengan performansi yang kurang baik. Maka dari itu, sebaiknya pengelola program studi bisa menerima calon mahasiswa yang berasal dari perguruan tinggi terakreditasi A. Variabel Linearitas Jurusan Statistika ternyata tidak berpengaruh signifikan. Banyak dari mahasiswa Magister Statistika ITS yang jenjang S1-nya tidak linear dengan jurusan Statistika, bisa dari Matematika, Pendidikan Matematika, maupun lainnya. Namun hal tersebut tidak mempengaruhi hasil performansi mahasiswa di akhir perkuliahan.

Variabel IPK S1 yang selama ini menjadi patokan utama pengelola program studi ternyata memang memiliki pengaruh yang signifikan terhadap performansi mahasiswa pada akhir masa kuliah. Untuk mahasiswa yang memiliki IPK dibawah 3 diprediksi memiliki performansi kinerja yang kurang baik dibandingkan dengan mahasiswa yang memiliki IPK diatas 3. Adapun variabel Lama Studi dan Waktu Tunggu Kuliah juga

berpengaruh. Semakin singkat waktu yang dibutuhkan untuk studi S1 maupun waktu tunggu kuliahnya, maka performansi mahasiswa akan diprediksi semakin baik. Hal ini sama dengan penelitian yang sebelumnya dilakukan oleh Shaziya, H dkk. (2015) yang mengatakan bahwa IPK S1, lama studi dan waktu tunggu memang berpengaruh signifikan terhadap hasil kinerja mahasiswa dalam menempuh program di jenjang berikutnya.

Variabel-variabel masukan (*input*) diatas dapat dijadikan pertimbangan untuk penerimaan mahasiswa baru bagi pengelola Program Studi Magister Statistika ITS untuk meningkatkan kualitas lulusan dari hasil level kinerja mahasiswanya. Dari analisis setiap variabel masukan tersebut, pengelola program studi dapat mengambil beberapa patokan pertimbangan untuk penilaian prediksi kinerja mahasiswa dengan beberapa batasan antara lain :

1. Mahasiswa yang berasal dari daerah Luar Jawa Timur dapat memiliki performansi yang lebih baik jika dibandingkan dengan mahasiswa yang berasal dari daerah Jawa Timur.
2. Performansi mahasiswa yang memiliki status Sudah Bekerja lebih baik dari pada mahasiswa yang belum bekerja.
3. Jalur penerimaan mahasiswa melalui Biaya Sendiri dan *Joint Degree* dapat meningkatkan performansi kinerja mahasiswa dibandingkan jalur penerimaan yang lainnya.
4. Batasan nilai Toefl untuk performansi mahasiswa yang baik adalah kurang dari 388.
5. Batasan nilai Tes Potensi Akademik (TPA) untuk performansi mahasiswa yang baik adalah kurang dari 391.
6. Mahasiswa yang berasal dari institusi ter-akreditasi A cenderung memiliki level performansi yang lebih baik dibandingkan dengan mahasiswa yang berasal dari institusi ter-Akreditasi B, C maupun belum ter-akreditasi.
7. Mahasiswa yang memiliki Indeks Prestasi Kumulatif (IPK) S1 kurang dari 3.00 cenderung memiliki performansi mahasiswa yang kurang baik.
8. Mahasiswa yang memerlukan lama studi lebih dari 8 semester ketika menempuh S1 memiliki level kinerja performansi tidak lebih baik dibandingkan mahasiswa dengan lama studi 7 atau 8 semester.

9. Mahasiswa yang memerlukan waktu tunggu untuk lanjut ke jenjang Magister selama lebih dari 3 semester cenderung lulus dengan performansi kinerja yang menurun.

#### **4.5.4. Analisis Manfaat untuk Pengelola Program Studi**

Pada bagian ini, penulis akan memaparkan manfaat-manfaat yang dapat digunakan dengan sebaik-baiknya oleh pengelola program studi, dalam hal ini Program Studi Magister Statistika ITS, dengan adanya penelitian prediksi kinerja mahasiswa. Dengan adanya prediksi kinerja mahasiswa, pengelola program studi Magister Statistika ITS dapat melakukan evaluasi sebagai salah satu bahan pertimbangan penerimaan mahasiswa baru di lingkungan Magister Statistika ITS. Dengan data-data calon mahasiswa pada masa studi sebelumnya, kinerja dari calon mahasiswa dapat diprediksi dengan baik. Prediksi kinerja tersebut dapat dijadikan salah satu pertimbangan penerimaan mahasiswa baru. Pengelola program studi juga dapat melakukan evaluasi untuk peningkatan kinerja yang lebih baik bagi mahasiswa. Adapun hasil analisis pengaruh variabel yang berpengaruh signifikan di penelitian ini dapat dijadikan patokan data guna pertimbangan penerimaan mahasiswa baru di masa mendatang.

Adanya level performansi sebagai output dari prediksi kinerja mahasiswa membantu pengelola program studi untuk mendeteksi secara dini bagaimana performansi mahasiswa di masa mendatang dalam menempuh jenjang Magister di Statistika ITS. Deteksi dini tersebut bermanfaat untuk pencegahan awal bagi mahasiswa yang memiliki prediksi performansi yang kurang baik sehingga dapat ditingkatkan performansinya menjadi lebih baik lagi.

#### **4.6. Model Prediksi Kinerja Mahasiswa menggunakan SVM**

Setelah berbagai macam uji coba beserta analisisnya, sesuai dengan tujuan awal penelitian, penulis akan memaparkan model prediksi kinerja mahasiswa menggunakan *Support Vector Machine* yang ideal untuk digunakan sebagai evaluasi bagi Program Studi Magister Statistika ITS. Sistem prediksi dibangun dengan penggunaan metode SVM dengan fungsi *Radial Basis Function* (RBF) Kernel yang memiliki sigma parameter RBF sebesar 3,75. Penggunaan parameter tersebut merupakan parameter yang paling baik dilihat dari Akurasi, Presisi dan *Recall* untuk mendapatkan TP, FP dan FN yang paling optimal. Prediksi kinerja mahasiswa Magister Statistika ITS yang paling

baik dapat dibangun dengan 9 (sembilan) variabel masukan (*input*) data mahasiswa yang berpengaruh signifikan antara lain Asal Daerah, Status Kerja, Jalur Masuk, Nilai Toefl, Nilai TPA, Akreditasi institusi sebelumnya, IPK S1, Masa Studi S1 dan Waktu tunggu studi. Dari variabel masukan tersebut, akan diproses oleh sistem hingga muncul variabel keluaran (*output*) berupa Level Kinerja Mahasiswa dari level 0 hingga level 7 dengan kelompok performansi kinerja Kurang Baik dan Baik.

## **BAB 5**

### **PENUTUP**

Bab ini menguraikan beberapa hal yang menyatakan kesimpulan dari penelitian Prediksi Kinerja Mahasiswa menggunakan *Support Vector Machine* (SVM) untuk Pengelola Program Studi di Perguruan Tinggi yang dibuat setelah dilakukan pengujian dan analisa. Adapun terdapat saran-saran yang mendukung dalam upaya untuk lebih menyempurnakan penelitian-penelitian ke depannya.

#### **5.1. Kesimpulan**

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan, antara lain:

1. Model prediksi kinerja mahasiswa berhasil dibangun menggunakan metode *Support Vector Machine* dengan beberapa *input* variabel data yang kemudian dapat menghasilkan prediksi level performansi mahasiswa tersebut.
2. *Support Vector Machine* (SVM) dapat digunakan sebagai salah satu metode prediksi kinerja mahasiswa dengan data mahasiswa Tahun 2011-2015 yang menghasilkan prosentase akurasi sebesar 41,49%, presisi sebesar 79,34% dan *recall* sebesar 46,68%.
3. Ketika pengujian dilakukan dengan pengelompokan level performansi dengan jenis kelompok Kurang Baik dan Cukup Baik, nilai ketepatan klasifikasi performansi mahasiswa dapat meningkat dengan prosentase akurasi sebesar 74,84%, presisi sebesar 81,23%, dan *recall* sebesar 90,49%.
4. Dalam ketepatan prediksi, terdapat lebih banyak data mahasiswa yang memiliki hasil kesalahan prediksi bersifat positif (*false positive*), dan hanya sedikit mahasiswa yang memiliki kesalahan prediksi bersifat negatif (*false negative*). Artinya, sistem yang dibangun dengan metode SVM sudah cukup baik dalam memberikan deteksi dini sebagai *warning* bagi pengelola program studi mengenai kinerja mahasiswanya.
5. Fungsi Kernel yang paling optimal untuk digunakan sebagai pendamping metode SVM yang dapat menghasilkan nilai akurasi paling baik adalah fungsi *Linear Kernel* dengan nilai parameter C sebesar 5,00.

6. Optimasi Fungsi Kernel juga dilakukan untuk mendapatkan hasil dengan *false positive* yang lebih baik daripada *false negative*. Adapun fungsi kernel paling optimal untuk uji coba tersebut adalah *Radial Basis Function* (RBF) Kernel dengan parameter sigma RBF sebesar 1,75.
7. Dalam penelitian Prediksi Kinerja Mahasiswa di Program Studi Statistika ITS ini, metode SVM memiliki performa ketepatan prediksi yang lebih baik jika dibandingkan dengan metode statistik konvensional, Regresi Logistik yang hanya bisa menghasilkan akurasi 32,31%, presisi 50,35% dan *recall* 48,74%.
8. Terdapat indikasi bahwa banyak mahasiswa yang berpotensi berada pada level performansi yang tinggi, hanya saja karena satu dan lain hal maka mahasiswa tersebut cukup puas berada pada level performansi yang kurang maksimal.
9. Variabel masukan (*input*) yang berpengaruh signifikan terhadap prediksi kinerja mahasiswa antara lain Status Kerja, Jalur Masuk, Nilai Toefl, Nilai TPA, Akreditasi Asal Institusi, Indeks Prestasi Kumulatif S1, Lama Studi dan Waktu Tunggu. Adapun variabel Jenis Kelamin, Usia, Asal Kota dan Linieritas Jurusan tidak berpengaruh signifikan terhadap variabel respon.
10. Pengelola program studi dapat memaksimalkan kinerja performansi mahasiswanya dengan men-*filter* mahasiswa dengan beberapa kriteria berdasarkan variabel masukan (*input*) yang signifikan.

## 5.2. Saran

Adapun saran yang ingin disampaikan penulis untuk pengembangan lebih lanjut dari penelitian ini antara lain:

1. Pengembangan penelitian yang akan datang perlu dilakukan dengan dataset yang relatif besar agar tingkat keakuratan performansi model terjamin.
2. Perlu dilakukan penelitian lebih lanjut mengenai metode pemilihan variabel prediktor yang lebih detail dan signifikan, serta penentuan parameter SVM yang paling optimal.
3. Klasifikasi menggunakan metode *Support Vector Machine* yang sudah memiliki hasil baik perlu dikombinasikan dengan penggunaan Regresi Logistik untuk mendapatkan hasil implementasi yang lebih maksimal. Kombinasi tersebut dinamakan *Support Vector Regression*.

## DAFTAR PUSTAKA

- A. Karatzoglou, A. Smola, K. Hornik, and Zeieis, A. (2004) *kernelab – An S4 Package for Kernel Methods in R*. Journal of Statistical Software, vol. 11(9), pp. 1-20.
- Al-Barrak, M. A. dan Al-Razgan, M. (2016) *Predicting Students Final GPA Using Decision Trees: A Case Study*. International Journal of Information and Education Technology, Vol. 6, No. 7.
- Asogwa, O. C., dan Oladugba, A. V. (2015) *Of Students Academic Performance Rates Using Artificial Neural Networks (ANNs)*. American Journal of Applied Mathematics and Statistics.
- Barthos, H. Basir. (1992) *Perguruan Tinggi Swasta di Indonesia: Proses Pendirian Penyelenggaraan dan Ujian*. Jakarta: Bumi Aksara.
- Bolon-Canedo, V., Sanchez-Marono, N., Alonso-Betanzos, A., Benitez, J. M., Herrera, F. (2014) *A Review of Microarray Datasets and Applied Feature Selection Methods*. Information Sciences. 282: 111–135.
- Cristianini dan Taylor, S. (2000) *An introduction to Support Vector Machines*. Cambridge University Press.
- El-Halees, A. (2009) *Mining Students Data to Analyze Learning Behavior: A Case Study*. Tunisia: Conference Proceedings, University of Sfax, Tunisia.
- Feldman, D. C. (2004) *Managing Individual Are Group. Behavioral in Organization*. New York: McGraw Hill.
- Goller, et.al. (2000) *Automatic Document Classification: A Thorough Evaluation of Various Methods*. USA: Proceedings of International Symposium on Information Theory and Its Application, pp. 145-162.
- Gunawan, Rapoport, disitasi Madya,2006:2007 *Metodelogi Tindakan Action Research*.
- Hastuti, K. (2012) *Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif*. Semarang: Seminar Nasional Teknologi Informasi & Komunikasi Terapan.

- Hsu, C. W., Chang, C. C., Lin, C. J. (2003) *A Practical Guide to Support Vector Classification*.
- Igiri, C., P. (2015) *Support Vector Machine–Based Prediction System for a Football Match Result*.
- Kamagi. 2014. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. Tangerang: Universitas Multimedia Nusantara.
- Kaur, P., Singh, M., Josan, G. S. (2015) *Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector*. Punjab Technical University India.
- Mustafa. (2014) Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIL Dipanegara Makassar). Makassar: Citec Jurnal Vol : 1. STMik Dipanegara.
- Mustakim. (2015) Pengembangan Aplikasi Prediksi Penyakit Berbahaya Di Provinsi Riau Berdasarkan Model Markov *Chains*. Pekanbaru: Jurnal Manajemen dan Rekayasa Sistem Informasi, Vol: 1 Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003) *Support Vector Machine dan Aplikasinya Dalam Bioinformatika*. Diakses 15 Januari 2017, dari <http://www.ilmukomputer.com>.
- Nurahayati, S., Kusriani dan Luthfi, E., T. (2015) Prediksi Mahasiswa *Drop Out* Menggunakan Metode *Support Vector Machine*.
- Parikh, K., S. dan Shah, T., P. (2016) *Support Vector Machine – a Large Margin Classifier to Diagnose Skin Illnesses*.
- Permana, R., A. (2016) Seleksi Atribut pada Metode *Support Vector Machine* untuk Menentukan Kelulusan Mahasiswa *E-Learning*.
- Prasetyo, Gusti R. A. dan Robandi, Imam. (2008) Peramalan Beban Jangka Pendek Untuk Hari-hari Libur Dengan Metode *Support Vector Machine*. Surabaya: Tugas Akhir, ITS.



- Rutoto, Sabar. (2007) Pengantar Metodologi Penelitian. FKIP: Universitas Muria Kudus.
- Sembiring, Krisantus. (2007) Tutorial SVM Bahasa Indonesia. Skripsi S1 Teknik Informatika, ITB. Bandung: Institut Teknologi Bandung.
- Santosa, Budi. (2007) Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- Scholkopf, B. dan Smola, A. (2002) *Learning with Kernels*. Cambridge, Massachusetts: The MIT Press.
- Shaziya, H., dkk. (2015) *Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier*.
- Sugiyono. (2011) Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: AFABETA, cv.
- Teshnizi, S. H., dkk. (2015) *A Comparison of Logistic Regression Model and Artificial Neural Networks in Predicting of Student's Academic Failure*.
- Theopilus, B S., Permanasari, A., E., dan Hidayah, I. (2014) Komparasi Kernel pada Algoritma *Support Vector Machine* Studi Kasus Klasifikasi Penjurusan di SMA Saverius Sragen.
- Vapnik, V, (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag.
- Zhang, Y dan Wang, W. (2010) *Pattern Classification of Electroencephalography from the Typical Specialized Students*. Education Technology and Computer Science (ETCS), International Workshop., vol. 1, pp. 836-839.

*halaman ini sengaja dikosongkan*

## BIOGRAFI PENULIS



**Nama** : Fathin Hilmiyah  
**Tempat Tanggal Lahir** : Surabaya, 15 Januari 1993  
**Email** : fathinhilmiyah@gmail.com

Penulis merupakan mahasiswa yang berasal dari Kota Surabaya, Jawa Timur. Penulis merupakan anak kedua dari dua bersaudara. Penulis menamatkan pendidikan dasar di SD Al-Falah Tropodo Sidoarjo pada tahun 2004, menamatkan pendidikan Sekolah Menengah Pertama di SMP Al-Falah Deltasari Sidoarjo pada tahun 2007 dan menamatkan pendidikan Sekolah Menengah Atas di SMAN 16 Surabaya pada tahun 2010. Setelah itu Penulis melanjutkan studi ke jenjang D3 Teknik Informatika di Politeknik Elektronika Negeri Surabaya (PENS-ITS) dan tamat pada tahun 2013. Pada tahun 2013 Penulis melanjutkan jenjang D4 Teknik Informatika PENS, dan lulus pada tahun 2015.

Setelah menamatkan jenjang pendidikan D4, Penulis melanjutkan pendidikan ke jenjang Magister (S2) pada tahun 2015 di Departemen Manajemen Teknologi Fakultas Bisnis dan Manajemen Teknologi Institut Teknologi Sepuluh Nopember Surabaya (MMT-ITS) dengan mengambil bidang konsentrasi/keahlian Manajemen Teknologi Informasi. Penulis memiliki ketertarikan pada bidang data mining.