

**KLASIFIKASI LAMA STUDI MAHASISWA FSM UNIVERSITAS  
DIPONEGORO MENGGUNAKAN REGRESI LOGISTIK BINER DAN  
*SUPPORT VECTOR MACHINE (SVM)***



---

---

**SKRIPSI**

---

---

**Disusun Oleh :**  
**SRI MAYA SARI DAMANIK**  
**24010210120002**

**JURUSAN STATISTIKA**  
**FAKULTAS SAINS DAN MATEMATIKA**  
**UNIVERSITAS DIPONEGORO**  
**SEMARANG**  
**2014**

**KLASIFIKASI LAMA STUDI MAHASISWA FSM UNIVERSITAS  
DIPONEGORO MENGGUNAKAN REGRESI LOGISTIK BINER DAN  
*SUPPORT VECTOR MACHINE (SVM)***

**Disusun oleh**

**SRI MAYA SARI DAMANIK**

**24010210120002**

**SKRIPSI**

**Diajukan Sebagai Syarat untuk Mendapatkan Gelar Sarjana  
pada Jurusan Statistika**

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2014**

## HALAMAN PENGESAHAN I

Judul Skripsi : Klasifikasi Lama Studi Mahasiswa FSM Universitas  
Diponegoro Menggunakan Regresi Logistik Biner dan  
*Support Vector Machine* (SVM)

Nama Mahasiswa : Sri Maya Sari Damanik

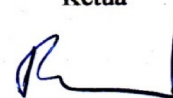
NIM : 24010210120002

Telah diujikan pada sidang Tugas Akhir tanggal 27 November 2014 dan  
dinyatakan lulus pada tanggal 4 Desember 2014

Semarang, Desember 2014

Panitia Penguji Ujian Tugas Akhir,

Ketua



Prof. Drs. Mustafid, M.Eng, Ph.D

NIP. 195505281980031002

Mengetahui,



Ketua Jurusan Statistika  
Fakultas Sains dan Matematika Undip

Dra. Dwi Ispriyanti, M. Si.

NIP. 195709141986032001

## HALAMAN PENGESAHAN II

Judul Skripsi : Klasifikasi Lama Studi Mahasiswa FSM Universitas  
Diponegoro Menggunakan Regresi Logistik Biner dan  
*Support Vector Machine (SVM)*

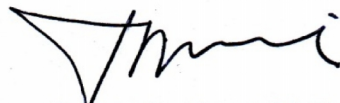
Nama Mahasiswa : Sri Maya Sari Damanik

NIM : 24010210120002

Telah diujikan pada sidang Tugas Akhir tanggal 27 November 2014

Semarang, Desember 2014

Pembimbing I



Dr. Dwi Ispriyanti, M. Si.  
NIP. 195709141986032001

Pembimbing II



Sugito, S.Si, M. Si.  
NIP. 197610192005011001

## KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus, karena kasih dan anugrahNya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul “**Klasifikasi Lama Studi Mahasiswa FSM Universitas Diponegoro Menggunakan Regresi Logistik Biner dan *Support Vector Machine* (SVM)**”.

Penulis menyadari laporan ini tidak akan dapat diselesaikan tanpa bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M. Si. selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro dan sekaligus dosen pembimbing I.
2. Bapak Sugito, S. Si, M. Si selaku dosen pembimbing II yang telah berkenan meluangkan waktu dalam memberikan masukan, arahan, dan bimbingan kepada penulis.
3. Seluruh Bapak dan Ibu Dosen Jurusan Statistika FSM Universitas Diponegoro yang telah memberikan ilmu yang sangat berguna.
4. Semua pihak yang telah membantu dalam penulisan laporan ini yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan. Oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan.

Semarang, Desember 2014

**Penulis**

## ABSTRAK

Wisuda adalah hasil akhir dari proses kegiatan belajar mengajar selama mengikuti perkuliahan di perguruan tinggi. Dalam mencapai gelar S1 membutuhkan waktu normal yaitu selama empat tahun, tetapi ada banyak mahasiswa yang menyelesaikan studinya melebihi batas normal (lebih dari empat tahun) dan ada juga yang kurang dari empat tahun. Lama studi mahasiswa dapat dipengaruhi oleh banyak faktor antara lain Indeks Prestasi Kelulusan (IPK), jenis kelamin, jurusan, lama studi yang ditempuh, beasiswa, *part time*, organisasi, dan jalur masuk universitas. Pada penelitian ini, akan dilakukan klasifikasi berdasarkan status lama studi mahasiswa lebih dari empat tahun dan kurang dari sama dengan empat tahun. Metode yang digunakan untuk klasifikasi lama studi mahasiswa dengan jenis data nominal adalah Metode *Support Vector Machine* (SVM) dan akan dibandingkan dengan metode Regresi Logistik Biner. Berdasarkan hasil penelitian dengan metode regresi logistik biner, menunjukkan variabel yang berpengaruh terhadap lama studi mahasiswa adalah Jurusan dan IPK dengan ketepatan klasifikasi 70%. Sedangkan ketepatan klasifikasi dengan menggunakan SVM ketepatan klasifikasi tertinggi dengan menggunakan kernel linear, *Polynomial* dan RBF mencapai 90%.

**Kata kunci** : Lama studi, Regresi Logistik Biner, *Support Vector Machine* (SVM), Ketepatan Klasifikasi.

## ABSTRACT

Graduation is the end result of the process of learning for studying in college. In attaining a normal S1 takes that for four years, but there are plenty of students who completed their studies beyond normal limits (over four years) and there is also less than four years. Older students study can be affected by many factors, among others, Graduation Achievement Index (GPA), gender, major, long pursued studies, scholarships, part-time, organizations, and university entrance pathway. In this study, the classification will be based on the status of a student's study time more than four years and less than or equal to four years. The method used to study the old classification of students with nominal data type is the method of Support Vector Machine (SVM) and will be compared with the Binary Logistic Regression method. Based on the research results of the binary logistic regression method, showed variables influencing the study period students are Subject and GPA with a classification accuracy of 70%. While the classification accuracy by using SVM highest classification accuracy using a linear kernel, polynomial and RBF reached 90%.

**Keywords:** Older studies, Binary Logistic Regression, Support Vector Machine (SVM), Classification Accuracy.

## DAFTAR ISI

|  | Halaman |
|--|---------|
| HALAMAN JUDUL .....                                | i       |
| HALAMAN PENGESAHAN I .....                         | ii      |
| HALAMAN PENGESAHAN II .....                        | iii     |
| KATA PENGANTAR .....                               | iv      |
| ABSTRAK .....                                      | v       |
| ABSTRACT .....                                     | vi      |
| DAFTAR ISI .....                                   | vii     |
| DAFTAR TABEL .....                                 | x       |
| DAFTAR GAMBAR .....                                | xi      |
| DAFTAR LAMPIRAN .....                              | xii     |
| <b>BAB I     PENDAHULUAN</b>                       |         |
| 1.1. Latar Belakang .....                          | 1       |
| 1.2. Permasalahan .....                            | 3       |
| 1.3. Batasan Permasalahan .....                    | 4       |
| 1.4. Tujuan .....                                  | 4       |
| <b>BAB II    TINJAUAN PUSTAKA</b>                  |         |
| 2.1. Fakultas Sains dan Matematika (FSM) .....     | 5       |
| 2.1.1. Sejarah Fakultas Sains dan Matematika ..... | 5       |
| 2.1.2. Visi dan Misi FSM UNDIP .....               | 8       |
| 2.1.3. Tujuan FSM UNDIP .....                      | 9       |
| 2.2. Peraturan Akademik .....                      | 9       |

|         |  |    |
|---------|--|----|
| 2.3.    | Klasifikasi.....   | 10 |
| 2.4.    | Regresi Logistik Biner.....  | 12 |
| 2.4.1.  | Estimasi Parameter.....  | 13 |
| 2.4.2.  | Uji Signifikansi.....  | 16 |
| 2.5.    | <i>Support Vector Machine</i> (SVM).....                                   | 18 |
| 2.5.1.  | Klasifikasi Linear <i>Separable</i> .....                                  | 19 |
| 2.5.2.  | Klasifikasi Linear <i>Non-Separable</i> .....                              | 22 |
| 2.5.3.  | Klasifikasi Non-Linear.....  | 24 |
| <br>    |  |    |
| BAB III | METODOLOGI PENELITIAN  |    |
| 3.1.    | Lokasi Penelitian.....   | 26 |
| 3.2.    | Jenis dan Sumber Data.....   | 26 |
| 3.3.    | Populasi dan Sampel.....   | 27 |
| 3.4.    | Variabel Penelitian.....   | 27 |
| 3.5.    | Langkah Analisis Data.....   | 28 |
| 3.6.    | Diagram Alir Analisis Data.....  | 30 |
| <br>    |  |    |
| BAB IV  | HASIL DAN PEMBAHASAN   |    |
| 4.1.    | Analisis Deskriptif.....   | 31 |
| 4.2.    | Analisis Lama Studi Mahasiswa dengan Metode Regresi<br>Logistik Biner..... | 38 |
| 4.2.1.  | Model Awal.....  | 38 |
| 4.2.2.  | Uji Rasio Likelihood.....  | 39 |
| 4.2.3.  | Uji Wald.....  | 40 |
| 4.2.4.  | Uji Rasio Likelihood Kedua.....  | 43 |
| 4.2.5.  | Uji Wald kedua.....  | 43 |

|        |   |    |
|--------|---|----|
| 4.2.6. | Uji Kesesuaian Model.....   | 45 |
| 4.2.7. | Model Akhir.....  | 45 |
| 4.2.8. | Interpretasi Model Regresi Logistik Biner.....  | 46 |
| 4.2.9. | Ketepatan Klasifikasi Menggunakan Regresi Logistik<br>Biner.....  | 47 |
| 4.3.   | Klasifikasi Lama Studi Mahasiswa dengan Metode <i>Support<br/>Vector Machine</i> (SVM).....                               | 48 |
| 4.3.1. | Klasifikasi SVM Menggunakan Fungsi Kernel Linier.<br>.....  | 48 |
| 4.3.2. | Klasifikasi SVM Menggunakan Fungsi Kernel<br><i>Polynomial</i> .....  | 49 |
| 4.3.3. | Klasifikasi SVM Menggunakan Fungsi Kernel RBF...<br>.....   | 51 |
| 4.4.   | Perbandingan Ketepatan Klasifikasi dengan Metode Regresi<br>Logistik Biner dan Metode <i>Support Vector Machine</i> ..... | 54 |
| BAB V  | KESIMPULAN .....  | 55 |
|        | DAFTAR PUSTAKA .....  | 56 |
|        | LAMPIRAN.....   | 58 |

## DAFTAR TABEL

|  | Halaman |
|--|---------|
| Tabel 2.1. Matriks Konfusi untuk klasifikasi dua kelas.....                                    | 11      |
| Tabel 2.2. Fungsi Kernel yang dipakai dalam penelitian.....                                    | 25      |
| Tabel 3.1. Variabel Penelitian.....  | 28      |
| Tabel 4.1. Ringkasan Data .....  | 31      |
| Tabel 4.2. Hasil Estimasi Parameter .....  | 39      |
| Tabel 4.3. Hasil Uji Wald .....  | 41      |
| Tabel 4.4. Hasil Uji Wald Kedua.....   | 44      |
| Tabel 4.5. Nilai Odds Ratio .....  | 46      |
| Tabel 4.6. Ketepatan Klasifikasi Regresi Logistik Biner.....                                   | 47      |
| Tabel 4.7. Nilai Error klasifikasi dengan Menggunakan Fungsi Kernel Linier ....                | 48      |
| Tabel 4.8. Matriks Konfusi dengan Menggunakan Fungsi Kernel Linier .....                       | 49      |
| Tabel 4.9. Nilai Error klasifikasi dengan Menggunakan Fungsi Kernel <i>Polynomial</i><br>..... | 50      |
| Tabel 4.10. Matriks Konfusi dengan Menggunakan Fungsi Kernel <i>Polynomial</i>                 | 51      |
| Tabel 4.11. Nilai Error Klasifikasi dengan Menggunakan Fungsi Kernel RBF ....                  | 52      |
| Tabel 4.12. Matriks Konfusi dengan Menggunakan Fungsi Kernel RBF .....                         | 53      |
| Tabel 4.13. Hasil Klasifikasi data <i>testing</i> dengan menggunakan SVM.....                  | 53      |
| Tabel 4.14. Perbandingan Klasifikasi dengan Regresi Logistik Biner dan SVM.                    | 54      |

## DAFTAR GAMBAR

|  | Halaman |
|--|---------|
| Gambar 2.1. Discrimination boundaries dan konsep fungsi pemisah.....         | 19      |
| Gambar 3.1. Diagram Alir Pengolahan Data .....                               | 30      |
| Gambar 4.1. Deskripsi Variabel Lama studi terhadap Variabel Jenis Kelamin .. | 32      |
| Gambar 4.2. Deskripsi Variabel Lama studi terhadap Variabel Jurusan .....    | 33      |
| Gambar 4.3. Deskripsi Variabel Lama studi terhadap Variabel IPK.....         | 34      |
| Gambar 4.4. Deskripsi Variabel Lama studi terhadap Variabel Beasiswa .....   | 35      |
| Gambar 4.5. Deskripsi Variabel Lama studi terhadap Variabel Parttime.....    | 36      |
| Gambar 4.6. Deskripsi Variabel Lama studi terhadap Variabel Organisasi ..... | 37      |
| Gambar 4.7. Deskripsi Variabel Lama studi terhadap Variabel Jalur Masuk..... | 38      |

## DAFTAR LAMPIRAN

|  | Halaman |
|--|---------|
| Lampiran 1. Data Lama Studi Mahasiswa Fakultas Sains dan Matematika Periode ke-131 sampai Periode ke-134 ..... | 52      |
| Lampiran 2. Tabulasi silang Variabel Prediktor terhadap Variabel Respon.....                                   | 58      |
| Lampiran 3. Output Regresi Logistik Biner menggunakan data training dan testing 90%:10%.....                   | 60      |
| Lampiran 4. Output Regresi Logistik Biner dengan menggunakan Variabel yang signifikan.....                     | 65      |
| Lampiran 5. Output Ketepatan Klasifikasi dengan Metode Regresi Logistik Biner .....                            | 69      |
| Lampiran 6. Output Klasifikasi SVM dengan Menggunakan data training dan testing 90%:10% .....                  | 70      |
| Lampiran 7. Syntax SVM .....   | 81      |
| Lampiran 8. Tabel Chi-Square.....  | 85      |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pendidikan Tinggi merupakan jenjang pendidikan setelah pendidikan menengah yang mencakup program pendidikan diploma, sarjana, magister, spesialis dan doktor yang diselenggarakan oleh Perguruan Tinggi. Kelulusan atau wisuda adalah hasil akhir dari proses kegiatan belajar mengajar selama mengikuti perkuliahan di perguruan tinggi. Universitas Diponegoro (UNDIP) adalah salah satu Universitas di Indonesia yang memiliki 11 (sebelas) fakultas. Fakultas Sains dan Matematika (FSM) adalah salah satu dari 11 Fakultas di Undip. FSM terdiri dari 7 (tujuh) jurusan dengan 6 (enam) program S1 yaitu Matematika, Biologi, Kimia, Fisika, Statistika, dan Teknik Informatika dan D3 Instrumentasi dan Elektronika. Lama Studi Program Sarjana menurut Peraturan Akademik UNDIP dijadwalkan dalam 8 semester (4 tahun) atau dapat ditempuh kurang dari 8 semester (4 tahun) dan selama-lamanya 14 semester (7 tahun). Setiap tahun UNDIP menyelenggarakan upacara wisuda dalam 4 periode yaitu periode Januari, April, Agustus, dan Oktober. Dalam 4 periode kelulusan jumlah lulusan dengan jumlah mahasiswa baru tidak sebanding.

Lama studi mahasiswa kemungkinan dapat dipengaruhi oleh banyak faktor. Faktor-faktor yang kemungkinan mempengaruhi dalam hal kelulusan antara lain Indeks Prestasi Kelulusan (IPK), jenis kelamin, jurusan, lama studi yang ditempuh,

beasiswa, *part time*, organisasi, dan jalur masuk universitas. Misalnya pada pendaftaran sidang yang sudah dijadwalkan, apabila mahasiswa tersebut terlambat sehari melakukan pendaftaran sidang dari batas yang ditentukan maka harus menunggu untuk sidang pada periode selanjutnya, dimana itu akan mempengaruhi lama studi mahasiswa tersebut. Mahasiswa yang belum merasa puas dengan IPK nya akan mengambil atau mengulang mata kuliah sampai memenuhi batas IPK yang diinginkan mahasiswa tersebut, yang dimana itu juga akan mempengaruhi lama studi. Oleh karena itu, peneliti ingin meneliti faktor-faktor yang mempengaruhi lama studi mahasiswa serta ingin mengklasifikasikan kelulusan mahasiswa ke dalam dua kategori yaitu lulus tepat waktu untuk mahasiswa yang menempuh pendidikan S1 kurang dari sama dengan 4 tahun (8 semester) dan lulus tidak tepat waktu untuk mahasiswa yang menempuh pendidikan lebih dari 4 tahun (8 semester). Metode klasifikasi yang digunakan dalam metode penelitian ini adalah metode *Support Vektor Machine* (SVM) dan akan dibandingkan dengan metode Regresi Logistik Biner. Kedua metode ini memiliki persamaan jenis nilai variabel data yaitu pada variabel responnya dimana menggunakan nilai variabel nominal.

*Support Vektor Machine* (SVM) merupakan salah satu bagian dari Data Mining yang digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Menurut Criastinini dan Shawe (2000) dalam Supriyanto (2013), konsep SVM dapat dijelaskan dengan cara sederhana sebagai usaha untuk mencari fungsi pemisah (*hyperplane*) terbaik dari berbagai alternatif garis pemisah yang mungkin. Menurut Supriyanto (2013), prinsip dasar SVM adalah

klasifikasi yang bersifat linier, dan selanjutnya dikembangkan agar dapat bekerja pada problem *non-linier* dengan menggunakan fungsi kernel (fungsi yang memudahkan proses pengklasifikasian data).

Berdasarkan uraian latar belakang diatas dengan menggunakan data lama studi mahasiswa FSM Undip wisuda periode ke-131 sampai dengan periode ke-134 maka lama studi mahasiswa tersebut akan diklasifikasikan dengan menggunakan metode *Support Vector Machine* (SVM) dan Regresi Logistik Biner.

## **1.2 Permasalahan**

Berdasarkan latar belakang diatas, masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut :

1. Bagaimana model dan ketepatan klasifikasi lama studi mahasiswa FSM Undip periode ke-131 sampai dengan periode ke-134 dengan menggunakan regresi logistik biner ?
2. Bagaimana ketepatan klasifikasi lama studi mahasiswa FSM Undip periode ke-131 sampai dengan periode ke-134 dengan menggunakan metode *Support Vector Machine* (SVM) ?
3. Bagaimana perbandingan ketepatan klasifikasi metode regresi logistik biner dan SVM pada klasifikasi lama studi mahasiswa lulusan FSM Undip periode ke-131 sampai dengan periode ke-134 ?

### 1.3 Batasan Permasalahan

Berdasarkan latar belakang diatas, maka lingkup permasalahan yang akan dibahas adalah sebagai berikut :

1. Data yang digunakan adalah data mahasiswa lulusan pada periode ke-131 sampai dengan periode ke-134 dengan lingkup wilayah adalah Fakultas Sains dan Matematika Undip.
2. Responden pada penelitian ini adalah alumni mahasiswa FSM program S1.
3. Metode klasifikasi yang digunakan adalah metode Regresi Logistik Biner dan *Support Vector Machine* (SVM) dengan kelas klasifikasi pada kedua metode adalah biner.

### 1.4 Tujuan

Adapun tujuan dilakukannya penelitian dari tugas akhir ini adalah

1. Memperoleh model dan ketepatan klasifikasi lama studi mahasiswa lulusan FSM Undip periode ke-131 sampai dengan periode ke-134 dengan menggunakan metode Regresi Logistik Biner.
2. Memperoleh ketepatan klasifikasi lama studi mahasiswa lulusan FSM Undip periode ke-131 sampai dengan periode ke-134 dengan menggunakan metode *Support Vector Machine* (SVM).
3. Memperoleh perbandingan ketepatan klasifikasi Regresi Logistik Biner dan *Support Vector Machine* (SVM) pada data lama studi mahasiswa lulusan FSM Undip periode ke-131 sampai dengan periode ke-134.