

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/323409685>

# Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa

Article · April 2018

CITATIONS

7

READS

14,465

3 authors, including:



**Randy Cahya Wihandika**  
Brawijaya University

65 PUBLICATIONS 142 CITATIONS

[SEE PROFILE](#)



**Dian Eka Ratnawati**  
Brawijaya University

55 PUBLICATIONS 77 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Human Detection and Tracking [View project](#)



Face Understanding [View project](#)

## Implementasi Algoritme *Support Vector Machine* (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa

Arif Pratama<sup>1</sup>, Randy Cahya Wihandika<sup>2</sup>, Dian Eka Ratnawati<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>arif.ilang@gmail.com, <sup>2</sup>randicahya@ub.ac.id, <sup>3</sup>dian\_ilkom@ub.ac.id

### Abstrak

Lulus tepat waktu adalah keinginan semua mahasiswa. Pada kenyataannya tidak seperti yang diharapkan banyak mahasiswa yang lulus lebih dari 4 tahun. Sehingga diperlukan penerapan prediksi kelulusan mahasiswa yang dapat mengklasifikasikan data prediksi kelulusan berdasarkan parameter-parameter yang telah ditentukan. Karena itu diperlukan penerapan sistem cerdas untuk dapat mengklasifikasikan data prediksi kelulusan mahasiswa berdasarkan parameternya. Algoritme *Support Vector Machine* (SVM) mengklasifikasikan data menjadi 2 kelas menggunakan *kernel Gaussian RBF* dengan kombinasi nilai parameter  $\lambda = 0,5$ , konstanta  $\gamma = 0,01$ , dan  $\epsilon$  (*epsilon*) = 0,001 itermax = 100,  $c = 1$  dengan menggunakan data latih sebanyak 170 *dataset*. Penelitian ini menghasilkan rata-rata akurasi sebesar 80,55 %.

**Kata kunci:** prediksi ketepatan waktu kelulusan mahasiswa, implementasi, *support vector machine*

### Abstract

*Graduate on time is the desire of all students. In reality, not as expected many students who graduated more than four years. necessitating the application of predictive graduation students can classify graduation prediction data based on parameters that have been determined. Because it is necessary for the application of intelligent systems can classify graduation prediction data based on parameters. Algorithm Support Vector Machine (SVM) to classify the data into two classes using kernel Gaussian RBF with a combined value of parameter  $\lambda = 0,5$ , constant  $\gamma = 0,01$ , and  $\epsilon$  (*epsilon*) = 0,001 itermax = 100,  $c = 1$  by using training data as much as 170 datasets, this study resulted in an average accuracy of 80,55 %.*

**Keywords:** graduation prediction, implementation, *support vector machine*

## 1. PENDAHULUAN

Lulus tepat waktu adalah keinginan semua mahasiswa. Tidak hanya itu, lulus tepat waktu adalah keuntungan bagi dua pihak. Pertama, untuk pihak mahasiswa, mahasiswa akan mendapatkan pekerjaan dengan lebih mudah karena perusahaan cenderung mencari yang baru lulus. Kedua, untuk pihak universitas, karena seiring tepatnya waktu kelulusan mahasiswa, hal itu akan membantu memajukan kualitas universitas tersebut, seperti peningkatan akreditasi. Tetapi dalam praktiknya mahasiswa tidak selalu dapat menuntaskan studinya selama waktu normal yang telah ditentukan. Banyak faktor yang menyebabkan ketidak tepatan waktu kelulusan mahasiswa tersebut. Faktor-faktor tersebut dapat bersumber

dari faktor internal dan faktor eksternal. Sehingga diperlukan penerapan prediksi kelulusan mahasiswa yang dapat mengklasifikasikan data prediksi kelulusan secara tepat berdasarkan parameter-parameter yang telah ditentukan. Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naïve Bayes* dengan jumlah dataset 150 memiliki rata-rata akurasi maksimum sebesar 76%. Karena penelitian tersebut memberikan akurasi yang kurang mencukupi, maka pada penelitian ini dilakukan dengan algoritma yang berbeda yaitu *Support Vector Machine*.

## 2. KAJIAN PUSTAKA

### 2.1 Prediksi Kelulusan

Prediksi Kelulusan adalah hasil dari

kegiatan memprediksi atau meramal atau memperkirakan lulus apabila mahasiswa tersebut telah memenuhi semua persyaratan yang telah ditetapkan oleh akademik misalkan telah menempuh 144 sks, telah lulus mata kuliah wajib, tidak memiliki nilai D lebih dari 10% , lulus ujian *ic3* dan *toefl* sebagai persyaratan wisuda, lulus ujian skripsi.

## 2.2 Klasifikasi

Klasifikasi adalah aspek penting yang ada pada data mining. Teknik klasifikasi telah banyak digunakan di berbagai permasalahan dalam suatu penelitian. Klasifikasi merupakan suatu metode pengelompokan data yang akan mempelajari data latih dengan menggunakan *algoritme* pengklasifikasian. Adapun beberapa algoritma klasifikasi, antara lain *Bayesian Classification*, *K-Nearest Neighbor*, *Decision Tree Induction*, *Case-Based Reasoning*, *Genetic Algorithms*, dan *Support Vector Machine*.

## 2.3 Support Vector Machine

SVM merupakan salah satu metode terbaik yang bisa dipakai dalam permasalahan klasifikasi. Konsep SVM bermula dari masalah klasifikasi dua kelas sehingga membutuhkan *training set* positif dan negatif. SVM berusaha menemukan *hyperplane* (pemisah) terbaik untuk memisahkan ke dalam dua kelas dan memaksimalkan margin antara dua kelas tersebut. Pada beberapa kasus, data tidak bisa diklasifikasi menggunakan metode *linier SVM*, sehingga dikembangkan fungsi *kernel* untuk mengklasifikasikan data dalam bentuk *non-linier*. Pada penelitian menggunakan *kernel Polynomial Degree*. *Sequential Training* memiliki *algoritme* yang lebih sederhana dan waktu yang diperlukan lebih cepat. Adapun *algoritme Sequential Training* adalah sebagai berikut:

1. Menginisialisasi  $\alpha_i = 0$  dan parameter lain, misalnya  $\lambda = 0,5$ ,  $\gamma = 0,01$ ,  $c = 1$ , IterasiMax = 100, dan  $\varepsilon = 0,001$ . Kemudian menghitung *matriks Hessian* dapat dihitung dengan rumus persamaan (1).

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (1)$$

2. Mulai dari data ke  $i$  sampai  $j$ , hitung menggunakan persamaan (2).

$$\begin{aligned} \text{a. } E_i &= \sum_{j=1}^n \alpha_j D_{ij} \\ \text{b. } \delta\alpha_i &= \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \end{aligned} \quad (2)$$

$$\text{c. } \alpha_i = \alpha_i + \delta\alpha_i$$

3. Langkah 2 dilakukan terus-menerus hingga kondisi iterasi maksimum tercapai atau  $\max(|\delta\alpha_i|) < \varepsilon$

Selanjutnya didapatkan nilai *support vector* (SV),  $SV = (\text{Threshold SV})$ . Nilai *Threshold SV* didapatkan dari beberapa percobaan, biasanya digunakan *threshold* > 0. Kemudian dilakukan proses *testing* untuk mendapatkan keputusan di mana fungsi keputusan dapat di hitung dengan persamaan (3).

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (3)$$

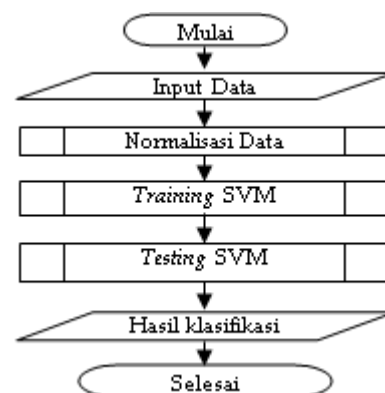
di mana nilai b dengan persamaan (4).

$$-\frac{1}{2} \left[ \sum_{i=1}^m \alpha_i y_i K(x_i, x^+) + \sum_{i=1}^m \alpha_i y_i K(x_i, x^-) \right] \quad (4)$$

## 3. PERANCANGAN

### 3.1. Diagram Alir

Gambar 1 merupakan diagram alir sistem secara keseluruhan. Proses diawali dengan input data gempa. Data tersebut dinormalisasi dengan menggunakan *Min-Max Normalization*. Data hasil normalisasi akan di *training* dan selanjutnya di *testing* dengan SVM, sehingga didapatkan hasil klasifikasi.



Gambar 1 Diagram Alir Sistem

## 4. PENGUJIAN DAN ANALISIS

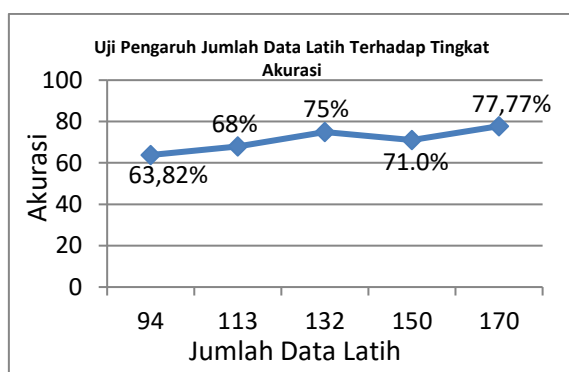
Pada pengujian implementasi *algoritme*

SVM pada data prediksi ketepatan waktu kelulusan mahasiswa ini menggunakan *dataset* dari bagian akademik Filkom yang berjumlah 188 data. Terdapat beberapa macam pengujian, yaitu:

1. Pengujian menggunakan jumlah data latih yang berbeda.
2. Pengujian pengaruh nilai variabel atau parameter pada *Sequential Training* SVM.
3. Pengujian pengaruh penggunaan *kernel*.

Pada pengujian kali ini menggunakan data latih perbandingan yang berbeda. Jumlah data latih yang digunakan antara lain 94, 113, 132, 150, dan 170 data. Sedangkan jumlah data uji yang digunakan untuk perbandingan antara lain 94, 75, 56, 38, dan 18 data. Nilai parameter *Sequential Training* yang digunakan tetap yaitu  $\lambda = 0,5$ ,  $\gamma = 0,01$ ,  $c = 1$ , IterasiMax = 100, dan  $\varepsilon = 0,0001$ , dan menggunakan *kernel* RBF.

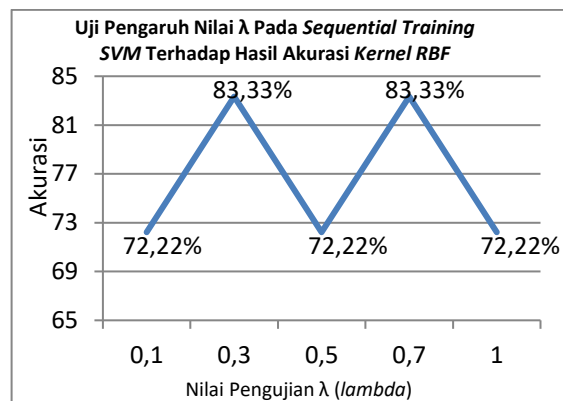
Pengujian menggunakan jumlah data latih yang berbeda yang terdapat pada gambar 2.



**Gambar 2** Grafik tingkat akurasi hasil pengujian jumlah data latih

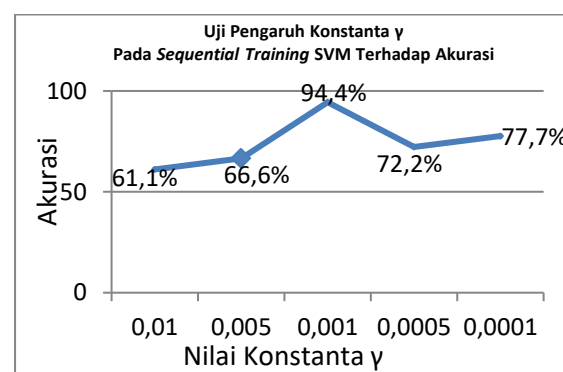
Berdasarkan skenario uji coba yang dilakukan pada semua data latih yang berbeda-beda tersebut didapatkan hasil akurasi yang berbeda-beda.

Pengujian pengaruh nilai variabel atau parameter pada *Sequential Training* SVM yang terdapat pada gambar 3.



**Gambar 3** Grafik tingkat akurasi pengaruh nilai  $\lambda$  (lambda)

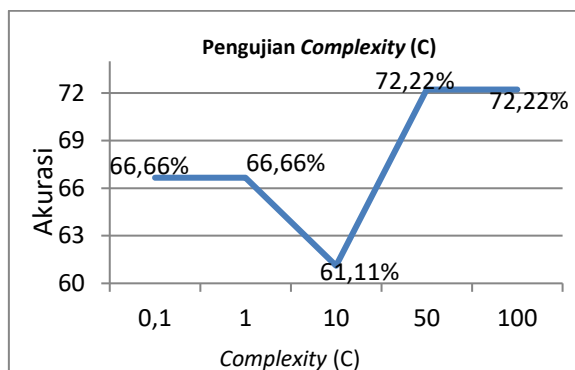
Pada analisa kali ini yaitu analisa nilai variabel atau parameter dari  $\lambda$  (lambda). Pada pengujian nilai parameter  $\lambda$  (lambda) yang telah dilakukan menghasilkan semakin besar nilai  $\lambda$  (lambda) maka akurasi yang dihasilkan cenderung berbeda-beda. Hal ini dapat disebabkan karena nilai  $\lambda$  (lambda) digunakan untuk menambah jarak atau *space* antar data sehingga dapat memberikan kemudahan *hyperplane* dalam membedakan kelas positif dan negatif yang terdapat pada gambar 4.



**Gambar 4** Grafik tingkat akurasi pengaruh konstanta  $\gamma$  (gamma)

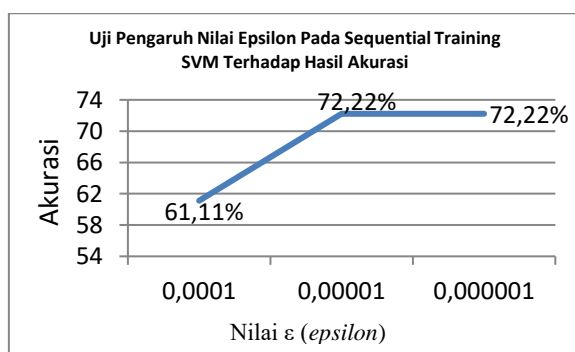
Pada pengujian parameter konstanta  $\gamma$  (gamma) didapatkan akurasi yang berbeda di mana akurasi yang menunjukkan penurunan didapatkan skenario 2 dan 5. Pada metode *sequential training* SVM ini, konstanta  $\gamma$  (gamma) digunakan untuk mengontrol kecepatan proses *learning* di mana kecepatan tersebut ditunjukkan oleh jumlah iterasi yang dibutuhkan untuk mencapai konvergensi. Hasil pengujian menunjukkan bahwa semakin kecil nilai konstanta  $\gamma$  (gamma) maka iterasi yang diperlukan semakin banyak dan dapat memberikan akurasi yang lebih baik. Hal ini disebabkan karena nilai konstanta  $\gamma$  (gamma)

mempengaruhi perhitungan  $\delta\alpha$ . Nilai  $\delta\alpha$  tersebut akan berkurang yang selisihnya tidak jauh dari nilai konstanta  $\gamma$  (*gamma*). Sehingga memerlukan proses iterasi lebih banyak untuk mencapai nilai  $\delta\alpha$  yang kurang dari  $\epsilon$  (*epsilon*). Tetapi dengan iterasi yang lebih banyak tersebut memberikan algoritme untuk dapat melakukan proses *learning* menjadi lebih baik. Sehingga dapat memberikan akurasi yang lebih baik di tunjukan pada gambar 5.



**Gambar 5** Grafik tingkat akurasi pengaruh nilai C

Pada pengujian parameter C didapatkan akurasi yang berbeda-beda pada iterasi yang berhenti berbeda. Pada metode *sequential training* SVM ini, nilai C digunakan untuk mengontrol *trade off* antara *margin* dan *error* klasifikasi. Tetapi pada pengujian yang telah dilakukan, nilai C tidak berpengaruh terhadap akurasi. Hal ini disebabkan karena data yang diuji dengan nilai C pada pengujian tersebut memiliki *error* klasifikasi yang berpengaruh pada *margin*. Sehingga akurasi yang didapatkan selalu berbeda-beda yang terdapat pada gambar 6.



**Gambar 6** Ditunjukkan bahwa grafik pengaruh nilai hasil akurasi  $\epsilon$  (*epsilon*).

Berdasarkan uji coba yang telah dilakukan ditunjukkan bahwa nilai variabel  $\epsilon$  (*epsilon*) yang semakin kecil menghasilkan akurasi yang semakin meningkat. Hal ini disebabkan karena

nilai variabel  $\epsilon$  (*epsilon*) yang semakin kecil dapat menghasilkan nilai  $\alpha$  yang optimal sehingga dapat menentukan *support vector* yang tepat. *Support vector* yang tepat tersebut dapat menghasilkan *hyperplane* (garis pemisah) untuk memisahkan kedua kelas tersebut semakin optimal ditunjukan pada tabel 1.

**Tabel 1** Hasil Pengujian jenis *kernel*.

Skenario	Gaussian RBF	Polynomial Degree	Linier
1	83.33%	77.77%	61.11%
2	77.77%	55.55%	72.22%
3	83.33%	61.11%	61.11%
4	77.77%	66.66%	77.77%
5	83.33%	61.11%	66.66%
6	77.77%	61.11%	55.55%
7	77.77%	77.77%	61.11%
8	77.77%	55.55%	50%
9	83.33%	61.11%	72.22%
10	83.33%	55.55%	66.66%
Rata-rata	80.55%	63.33%	64.44%

Pada tabel 1 menunjukkan bahwa *kernel RBF* menghasilkan akurasi yang baik. Dilihat dari rata-rata akurasinya *kernel RBF* lebih tinggi dari pada *kernel Polynomial* dan *Linier* karena memiliki rata-rata akurasi yaitu sebesar 80,55%. Hal ini disebabkan karena pada saat proses pemetaan data dalam dimensi yang lebih tinggi, *kernel RBF* merepresentasikan sebaran data yang lebih baik dibandingkan *kernel Polynomial* dan *Linier*.

## 5. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat diambil kesimpulan yaitu:

1. Rata-rata akurasi terbaik sebesar 80,55%. menggunakan *kernel Gaussian RBF* dengan kombinasi nilai variabel atau parameter  $\lambda = 0,5$ ,  $\gamma = 0,01$ ,  $C = 1$ , IterasiMax = 100, dan  $\epsilon$  SSS(*epsilon*) = 0,001 ,perbandingan data latih dan uji 170 dan 18.
2. Dari 3 jenis *kernel* yang digunakan yaitu *Polynomial Degree*, *Gaussian RBF*, dan *Linier* Pengaruh penggunaan *Kernel* pada *Algoritme Support Vector Machine* (SVM) yang paling baik yaitu menggunakan *kernel Gaussian RBF* karena *kernel* dengan rata-rata akurasi tertinggi adalah *kernel Gaussian RBF* yaitu 80,55% Hal ini Karena mengoptimasi nilai parameter  $\lambda$ ,  $\gamma$ ,  $C$ , IterasiMax, dan  $\epsilon$ (*epsilon*), pada *kernel SVM* untuk mendapatkan akurasi terbaik dan mengidentifikasi *kernel* yang terbaik

pula.

#### DAFTAR PUSTAKA

- Ramadhan, Dea Puspa, dkk. 2012. “Analisis Perbandingan *Opinion Mining* Berbahasa Indonesia Menggunakan *Support Vector Machine* dengan *Kernel Linear* dan *Radial Basic Function*”. Fakultas Informatika Institut Teknologi Telkom, Bandung.
- Fadilah, Arif Rahman. 2014. “Analisis dan Perbandingan Metode *Support Vector Machine* dan *Naive Bayes* Untuk *E-Mail Spam Filtering*”. Universitas Telkom. Bandung.
- Hardono, Jendi. 2013. “Klasifikasi Kondisi Penderita Penyakit Hepatitis dengan menggunakan *Support Vector Machine* (SVM)”, Skripsi, Universitas Brawijaya.
- Serli, Retno, 2013 “Pembangunan Model Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan *Neuro Fuzzy*.”
- Vijayakumar, Sethu and Si Wu. 1999, “*Sequential Support Vector Classifiers and Regression*”, Proceeding International Conference on Soft Computing (SOCO ‘99), Genoa, Italy, pp. 610-619.
- Khan, Aurangzeb, Baharum Baharudin, Lam Hong Lee, and Khairullah Khan. 2010. “A Review of *Machine Learning Algorithms for Text-Documents Classification*”, Journal of Advances in Information Technology, Vol. 1, No.1 hal 4-20.