

PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO

Suhardjono¹, Ganda Wijaya², Abdul Hamid³

Universitas Bina Sarana Informatika^{1,3}, STMIK Nusa Mandiri²
suhardjono@bsi.ac.id¹, ganda.gws@nusamandiri.ac.id², hamid.adh@bsi.ac.id³

ABSTRAKSI - Waktu kelulusan dengan tepat waktu bagi mahasiswa sangatlah penting untuk menentukan pekerjaan dalam perkuliahan, maka dari itu perlu di prediksi kelulusan mahasiswa sebelum akhir semester dengan menggunakan model support vector machine yang memiliki keuntungan dalam membuat data menjadi optimal tetapi support vector machine memiliki kekurangan dalam pengoptimal parameter. Particle swarm optimization dapat memperbaiki kekurangan yang terdapat pada support vector machine dalam hal mengoptimalkan parameter. Dari hasil yang didapat dengan menggunakan model support vector machine berbasis particle swarm optimization dapat meningkatkan akurasi prediksi dari sebesar 85.81% menjadi 86.43%. dengan kenaikan sebesar 00.62%. Sehingga dalam memprediksi kelulusan mahasiswa dapat akurat dan secara optimal dalam mengukur parameter yang diperlukan.

Kata kunci : Waktu Kelulusan, Particle Swarm Optimization, Support Vector Machine.

ABSTRACT - The time of graduation on time for students is very important to determine the work in lectures, so it is necessary to predict student graduation before the end of the semester by using a support vector machine model that has the advantage of making data optimal but the support vector machine has deficiencies in optimizing parameters. Particle swarm optimization can fix the deficiencies in vector machine support in terms of optimizing parameters. From the results obtained using the support vector machine model based on particle swarm optimization can increase the prediction accuracy from 85.81% to 86.43%. with an increase of 00.62%. So that in predicting student graduation can be accurate and optimally in measuring the required parameters.

Keywords: Graduation Time, Particle Swarm Optimization, Support Vector Machine

1. PENDAHULUAN

Kewajiban lembaga pendidikan menyediakan pendidikan berkualitas bagi anak didiknya. Keterlambatan atau waktu yang lama bagi anak didik mahasiswa untuk lulus mengakibatkan penurunan kualitas pendidikan dari perguruan tinggi tersebut.

Support Vector Machine (SVM) adalah sebuah metode seleksi yang membandingkan parameter standar seperangkat niali diskrit yang disebut kandidat set, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik (Dong, Xia, Tu, & Xing, 2007). SVM pertama kali diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992. SVM adalah salah satu teknik baru dibandingkan dengan teknik lain, tetapi memiliki performansi yang lebih baik di berbagai bidang aplikasi seperti bioinformatika, pengenalan tulisan tangan, klasifikasi teks, klasifikasi diagnosis penyakit dan lain sebagainya (Feng-Chia, 2009). Support Vector Machines (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi (Maimon, 2010). Dengan berorientasi pada tugas, kuat, sifat komputasi mudah dikerjakan, SVM telah mencapai sukses besar dan dianggap sebagai state-of-the-art classifier saat ini (Huang, Yang, King, & Lyu, 2008).

Karakteristik dari Support Vector Machine adalah sebagai berikut:

1. Support Vector Machine adalah linier classifier
2. Pattern Recognition dilakukan dengan mentransformasikan data pada input space ke ruang yang berdimensi lebih tinggi, dan optimasi dilakukan pada ruang vector yang baru.
3. Menerapkan strategi Structural Risk Minimization (SRM).
4. Prinsip kerja Support Vector Machine pada dasarnya hanya mampu menangani klasifikasi dua class.

SVM (Support Vector Machine) digunakan sebagai rangkaian harmonis konsep - konsep unggulan dalam bidang pattern recognition. Konsep dasar SVM merupakan salah satu metode yang terbimbing. Metode terbimbing yang dimaksud adalah metode yang membutuhkan data training dan data testing dalam uji coba. Menurut Basari et al (2013, 453) Support Vector Machine adalah metode untuk menganalisa data dan mengenali pola yang bisa digunakan untuk pengklasifikasian. Dalam pemakaian metode ini, harus memakai pelabelan. Pelabelan tersebut adalah kalimat positif atau kelas negatif.

Tujuan dari metode ini adalah menemukan hyperplane yang optimal yang memiliki margin

maksimal. Margin tersebut adalah jarak antara hyperplane dengan titik terdekat setiap kelas. Dewasa ini, SVM telah digunakan dalam masalah dunia nyata dan memberikan solusi yang lebih baik secara keseluruhan.

Pada umumnya masalah di dunia jarang yang bersifat linear, kebanyakan bersifat non linear. Agar dapat menyelesaikan masalah non linear, SVM menggunakan fungsi kernel. Fungsi kernel ini memetakan class yang ada pada input space berdimensi dua ke dalam vektor baru yang berdimensi tinggi.

Salah satu aspek penilaian kemampuan mahasiswa pada saat melamar kerja adalah waktu lulus mahasiswa tersebut selama menempuh pendidikan di Perguruan Tinggi tersebut

Mahasiswa dalam menempuh pendidikan tinggi dengan mengikuti proses belajar selama enam semester atau delapan semester akan mendapatkan sebuah nilai yang disebut indeks prestasi kumulatif (IPK). IPK bagi seorang mahasiswa sangatlah penting baik dalam segi belajar, mengikuti tugas akhir atau skripsi bahkan memasuki dunia pekerjaan. IPK juga sangat menentukan kelulusan mahasiswa, dan bagi perguruan tinggi kelulusan merupakan hal yang sangat penting dalam hal akreditasi. Maka memprediksi waktu kelulusan mahasiswa dalam prestasi akademik sangat penting bagi institusi pendidikan (Riyanto, Hamid, & Ridwansyah, 2019), dengan memprediksi waktu kelulusan tersebut dapat mengurangi resiko kegagalan kelulusan mahasiswa yang mengakibatkan rasio dosen di perguruan tinggi meningkat. Untuk mengurangi hal tersebut perlu dilakukan sebuah keputusan dalam hal kelulusan mahasiswa dengan pendekatan dan klasifikasi data mahasiswa yang sudah lulus.

Prediksi kelulusan mahasiswa pernah dilakukan dengan membandingkan ketiga model neural network (NN), decision tree(DT) dan support vector machine(SVM) dengan hasil akurasi tertinggi model SVM 85.18% (Riyanto et al., 2019). SVM telah berhasil diterapkan ke banyak bidang, mulai dari pengenalan wajah, pemrosesan data biologis untuk diagnosis medis, hingga prediksi waktu kelulusan (LU, MENG, & CAO, 2010). SVM dengan model supervised learning memiliki kelebihan dalam pengklasifikasian suatu data set sebagai titik dalam ruang vektor. Titik data baru didalam ruang vektor dipetakan kedalam ruang yang sama untuk diprediksi (Sugimoto, Takada, & Toi, 2013). SVM dengan kemampuan generalisasi yang berdimensi tinggi (Frias-Martinez, Sanchez, & Velez, 2010) untuk masalah nonlinear(Ren, 2012). dapat membuat pengklasifikasi data menjadi optimal (Xiang, 2013). Namun SVM sulit dalam pemilihan

parameter dengan sangat optimal (Ariyati, Ridwansyah, & Suhardjono, 2018).

Pemilihan parameter sering digunakan untuk memilih parameter yang optimal dalam menemukan parameter yang tidak relevan dan tidak optimal sehingga dapat dihilangkan parameter tersebut (Wang, Li, Song, Wei, & Lia, 2011), karena dapat mempengaruhi penurunan akurasi (Ridwansyah & Purwaningsih, 2018). Algoritma PSO dapat mengurangi permasalahan pemilihan parameter dengan memiliki kedua keuntungan yaitu menyederhanakan dan mengimplementasikan optimasi partikel asli dan keuntungan konvergensi cepat, mudah untuk keluar dari minimum lokal untuk algoritma optimasi. Particle Swarm Optimization (PSO) dapat menghasilkan perbaikan yang mengesankan dan lebih baik dalam menghadapi kelemahan pemilihan parameter yang dialami oleh model SVM (WU, 2007).

2. METODE PENELITIAN

Metode penelitian dapat diklasifikasikan berdasarkan tujuan dan tingkat kealamian (natural setting) obyek yang diteliti. Berdasarkan tujuan, metode penelitian dapat diklasifikasikan menjadi penelitian dasar (basic research), penelitian terapan (applied research) dan penelitian pengembangan (research and development). Selanjutnya berdasarkan tingkat kealamian, metode penelitian dapat dikelompokkan menjadi metode penelitian eksperimen, survey dan naturalistik.

Jenis penelitian eksperimen adalah metode yang menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkan dengan masalah penelitian. Jenis penelitian eksperimen dibagi dua, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor dosen terhadap kinerja. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing yang mana lebih baik

Teknik pengumpulan data adalah cara yang dapat digunakan untuk menggunakan data. Dalam pengumpulan data terdapat sumber yang terhimpun langsung oleh peneliti yang disebut data primer, sedangkan jika data diperoleh dari pihak kedua dapat disebut data sekunder.

Berbagai metode yang dilakukan dalam penelitian seperti SVM pernah dilakukan untuk memprediksi sesuatu objek. Dengan cara kerja SVM berdasarkan prinsip untuk menemukan hasil terbaik yang memisahkan dua kelas ruang input[10]. Penelitian yang dilakukan dengan metode SVM ini dimulai dengan pengumpulan data. Dataset diperoleh dari salah satu perguruan tinggi swasta dengan data

mahasiswa yang diambil dari tahun 2013-2018 sebanyak 796 paramater dengan sembilan atribut sebagai prediktor dan satu atribut sebagai hasil yaitu lulus tepat waktu atau tidak yang diuji coba menggunakan algoritma SVM dan SVM berbasis PSO untuk mengatasi masalah nilai yang rendah terhadap mahasiswa

dan untuk meningkatkan kinerja akademik siswa dalam dataset besar di mana kinerja siswa dalam ujian semester akhir dievaluasi. Pada Tabel 1 merupakan *sample* data yang kami kumpulkan dari salah satu perguruan tinggi swasta.

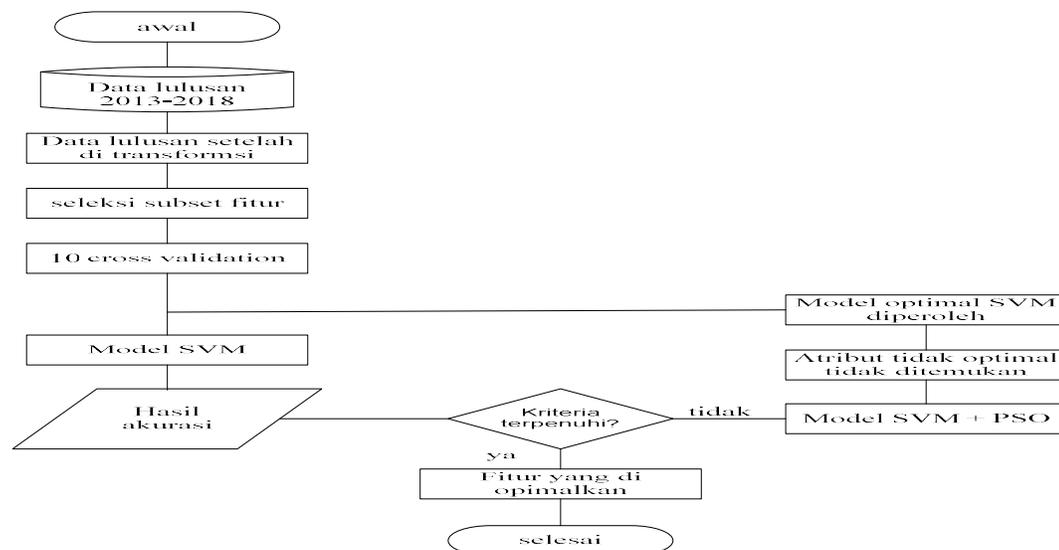
Tabel 1. Sample Data Kelulusan Mahasiswa

Jk	Jurusan SLTA	Asal SLTA	IPK1	IPK2	IPK3	IPK4	IPK 5	IPK 6	HASIL
L	IPA	SMA N 1 LAIS	2,55	2,61	2,64	2,74	2,97	3	YA
L	TKJ	SMK TI YPML	3,23	3,15	3,07	3,17	3,25	3,17	YA
P	AP	SMK Nusantara 1	2,73	2,73	2,62	2,64	2,89	3	YA
L	IPA	SMA Yuppentek 1	2,73	3,1	2,97	3,16	3,3	3,33	YA
L	IPS	Hidayatul Ikhwan	2,18	2,44	2,11	2,36	2,66	2,57	TIDAK
P	TKJ	SMK PUSTEK	2,32	2,39	2,43	2,74	2,67	2,95	TIDAK
P	PENJUALAN	SMK NEGERI 1	3,64	3,56	3,59	3,53	3,53	3,51	YA
L	OTOMOTIF	SMK	2,64	2,22	2,3	2,11	2,49	2,74	YA

Sebelum dilakukan klasifikasi data kami melakukan proses untuk menghilangkan data noise seperti data yang tidak lengkap dan hilang yang dapat memperlambat proses data saat klasifikasi. Setelah itu data di validasi dan diterapkan dengan metode SVM berbasis PSO yang dapat dilihat dari Gambar 2. Dengan algoritma yang kami usulkan akan dijelaskan secara rinci.

Metode yang diusulkan merupakan metode SVM berbasis PSO yang di tunjukan pada gambar 1 dengan mengolah data yang sudah bersih dari data noise dan data yang tidak lengkap yang di uji menggunakan metode SVM dengan menseleksi subset fitur.

Setelah menseleksi subset fitur akan dilakukan uji validasi dengan 10-cross-validation yang akan dilakukan dengan model SVM. Jika model di uji coba dan akan di validasi dengan mendapatkan hasil dari model SVM maka jika model kriteria tidak terpenuhi akan digunakan model SVM berbasis PSO yang setelah di uji coba akan menemukan atribut yang tidak optimal. Atribut yang tidak optimal akan digunakan model PSO untuk mengoptimalkan hasil akurasi. Jika menemukan fitur yang di optimalkan maka didapatkan hasil akurasi yang terbaik



Gambar 1. Model Penelitian yang diusulkan

3. HASIL DAN PEMBAHASAN

3.1. Support Vector Machine (SVM)

Setelah melakukan pengolahan data lulusan melalui software rapid miner dengan melakukan mentransformasi data untuk menseleksi subset fitur yang dilakukan dengan mencoba uji validasi 10 cross validation maka selanjutnya menggunakan model SVM dengan hasil yang didapat dilihat di tabel 2.

Tabel 2. Confusion Matrix SVM

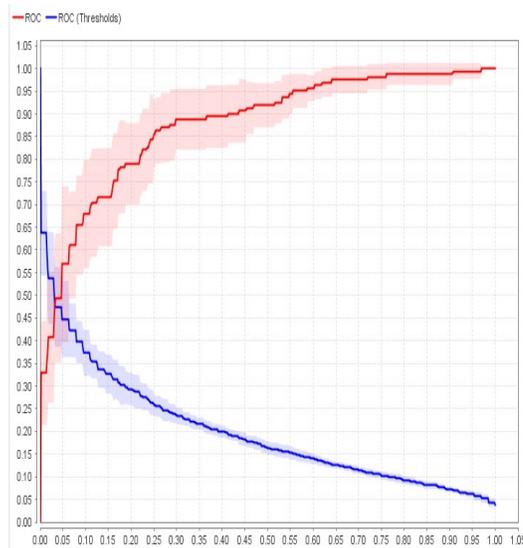
	Ya	Tidak
Prediksi Ya	615	94
Prediksi Tidak	19	68

Dari tabel 2 dapat dihitung dengan rumus akurasi sebagai berikut.

$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

$$\text{Akurasi} = \frac{(615 + 68)}{(615 + 19 + 68 + 94)} = 0.85804$$

didapat nilai akurasi SVM sebesar 0.85804 dan dibuat persen menjadi 85.81% dengan precision 79.33% dan recall 42.13%. Dari hasil tersebut didapat kernel model SVM dan Area Under Curve (AUC) yang dapat dilihat pada table 3 dan gambar 2.



Gambar 2. Area Under Curve (AUC) SVM

Dengan didapat kernel model SVM dari total parameter *support vector* 796 dengan mempunyai nilai bias sebesar -1.278 dan mendapatkan nilai bobot dari masing-masing atribut yang dapat dilihat dari tabel 3.

Tabel 3. Kernel Model SVM

Total number of Support Vectors	796
Bias (offset)	-1.278
w[JRS_SLTA]	0.048
w[ASAL_SLTA]	0.117
w[IPK1]	0.028
w[IPK2]	0.039
w[IPK3]	0.073
w[IPK4]	-0.324
w[IPK5]	-1.208
w[IPK6]	0.348

3.2. Support Vector Machine (SVM) berbasis Particle Swarm Optimization

Setelah dilakukan uji coba menggunakan model SVM maka akan dilakukan seleksi fitur untuk meningkatkan akurasi dari model SVM dengan menggunakan Particle Swarm Optimization (PSO), Sehingga nilai bobot dari masing-masing atribut akan maksimal.

Tabel 4. Bobot Atribut Model SVM Berbasis PSO

Atribut	Bobot
JRS_SLTA	0.897
ASAL_SLTA	1
IPK1	0
IPK2	0.832
IPK3	0
IPK4	0.399
IPK5	0.542
IPK6	0.089

Dari hasil tabel bobot atribut didapatkan 2 atribut yang tidak berpengaruh terhadap bobot nilai yaitu atribut ipk semester 1 (IPK1) dan ipk semester 3 (IPK3), sedangkan untuk atribut yang mempengaruhi bobot nilai ada 6 atribut yaitu atribut jurusan slta (JRS_SLTA), asal slt (ASAL_SLTA), ipk semester dua (IPK2), ipk semester empat (IPK4), ipk semester lima (IPK5) dan ipk semester enam (IPK6). Model SVM berbasis PSO dapat menghasilkan akurasi sebesar 86.43% dengan precision 83.78% dan recall 43.27%. Dari model SVM berbasis PSO akan menghasilkan nilai akurasi, rumus akurasi, kernel model dan kurva AUC.

Tabel 5. Confusion Matrix SVM berbasis PSO

	Ya	Tidak
Prediksi Ya	618	92
Prediksi Tidak	16	70

Berikut akurasi yang didapat dari tabel 5 dengan perhitungan rumus sebagai berikut

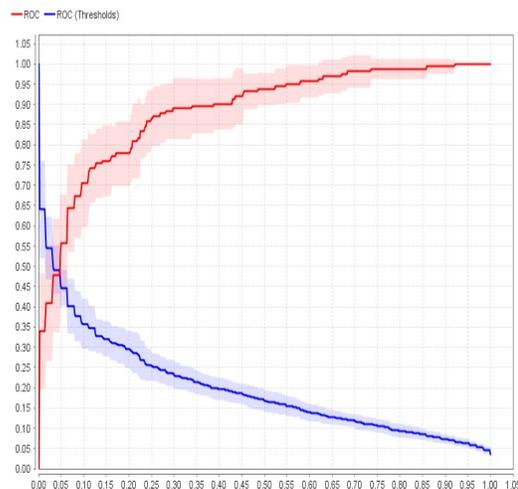
$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

$$\text{Akurasi} = \frac{(618 + 70)}{(618 + 16 + 70 + 92)} = 0.864322$$

didapat nilai akurasi SVM berbasis PSO sebesar 0.864322 dan dibuat persen menjadi 86.43% dengan precision 83.78% dan recall 43.27%.

Dari hasil tersebut didapat *Area Under Curve* (AUC) yang dapat dilihat pada gambar 3.

AUC: 0.884 +/- 0.038 (micro average: 0.884) (positive class: TIDAK)



Gambar 3. *Area Under Curve* (AUC) SVM Berbasis PSO

4. KESIMPULAN

Dari hasil eksperimen dengan menggunakan model support vector machine yaitu algoritma yang dalam pengklasifikasi data yang optima dapat menghasilkan akurasi sebesar 85.81% namun dengan akurasi seperti itu ada atribut yang tidak berpengaruh seperti atribut IPK semester 1 dan IPK semester 4 yang dalam hal ini tidak berbobot untuk menambah akurasi maka dari itu diperlukan pengoptimalan parameter dengan menggunakan particle swarm optimization yang mendapatkan akurasi sebesar 86.43%.

Dari model yang diusulkan dengan model particle swarm optimization dapat disimpulkan memiliki kenaikan sebesar 0.43% dalam meningkatkan akurasi dari model sebelumnya yaitu support vector machine. Maka algoritma support vector machine berbasis particle swarm optimization merupakan model yang terbaik dalam memprediksi kelulusan mahasiswa.

Penerapan metode SVM – PSO yang diwujudkan dalam prototipe aplikasi ketepatan waktu kelulusan mahasiswa akan sangat membantu untuk meningkatkan jumlah kelulusan ditahun yang akan datang. Sehingga

dapat meningkatkan kualitas lembaga pendidikan tersebut ke jenjang akreditasi yang lebih baik dari jenjang akreditasi sebelumnya

5. REFERENSI

- Ariyati, I., Ridwansyah, & Suhardjono. (2018). Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen. *JIKO (Jurnal Informatika dan Komputer) STMIK AKAKOM*, 3(2), 70–75. <https://ejournal.akakom.ac.id/index.php/jiko/article/view/127>
- Frias-Martinez, E., Sanchez, A., & Velez, J. (2010). Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition. *Engineering Applications of Artificial Intelligence*, 19(6), 693–704.
- LU, S.-X., MENG, J., & CAO, G.-E. (2010). SUPPORT VECTOR MACHINE BASED ON A NEW REDUCED SAMPLES METHOD. *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics*.
- Ren, J. (2012). ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *KnowledgeBased Systems*, 26, 144–153.
- Ridwansyah, & Purwaningsih, E. (2018). Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank. *Jurnal PILAR Nusa Mandiri*, 14(1), 83–88. <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/94>
- Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, 2(1). <http://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/6424>
- Sugimoto, M., Takada, M., & Toi, M. (2013). Comparison of robustness against missing values of alternative decision tree and multiple logistic regression for predicting clinical data in primary breast cancer. *Annual International Conference of the IEEE EMBS*.
- Wang, S., Li, D., Song, X., Wei, Y., & Lia, H. (2011). A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Syst. Appl*, 38(7), 8696–8702.
- WU, X. (2007). *Top 10 Algorithms in Data Mining*. New York: Springer.
- Xiang, C. (2013). A Chaotic Time Series Forecasting Model Based on Parameters Simultaneous Optimization Algorithm. *Journal of Information and Computational Science*, 10(15), 4917–4930