

# OPTIMASI KLASIFIKASI DATA KINERJA AKADEMIK MAHASISWA MENGGUNAKAN SVM BERBASIS ALGORITMA GENETIKA

Suamanda Ika Novichasari<sup>1</sup>, Wiwik Setyaningsih<sup>2</sup>

<sup>1,2</sup> Jurusan SI Teknik Informatika, Universitas Ngudi Waluyo  
Email : vichareal0311@gmail.com<sup>1</sup>,

**Abstract**— *For a college, especially a private university, students are the main component that supports the survival of the college. An educational database containing information about students is useful for predicting student academic performance. Several studies on the classification of academic performance have been conducted, it is clear that classification problems generally exist in the number of attributes, too many unnecessary attributes will increase computational time and reduce accuracy. The combination of PSO + SVM has proven to be more effective than SVM in various types of datasets. Therefore, this study will try to compare SVM-GA for the classification of student academic performance so that students who have good and bad academic performance can be seen. The data used is the academic performance data of the midwifery students of Ngudi Waluyo University, 2012-2014. The highest accuracy of SVM-GA is the accuracy of 93.55% and AUC 0.977. The previous SVM method had an accuracy of 90.51% and AUC 0.963. Based on the AUC value, the performance of the proposed SVM-GA method is in the "Perfect" group.*

**Keywords**— *Classification, academic performance, SVM, SVM-GA.*

## PENDAHULUAN

Institusi pendidikan memainkan peran penting bagi pertumbuhan dan perkembangan bangsa. Bagi sebuah perguruan tinggi khususnya perguruan tinggi swasta, mahasiswa merupakan komponen utama yang menunjang keberlangsungan hidup perguruan tinggi tersebut. Peraturan Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) NO 4 tahun 2017 tentang instrument akreditasi menyebutkan bahwa mahasiswa termasuk kriteria penilaian Akreditasi Program Studi (APS) dan Akreditasi Perguruan Tinggi (APT). Salah satunya keseimbangan rasio mahasiswa dengan dosen dan kebijakan penyelenggaraan system layanan bagi mahasiswa dalam menunjang proses pembelajaran yang efektif dan efisien. Tantangan terbesar institusi

pendidikan adalah menghadapi pertumbuhan data pendidikan yang eksplosif dan menggunakan data itu untuk meningkatkan kualitas keputusan manajerial. Database pendidikan yang berisi informasi tentang mahasiswa berguna untuk memprediksi kinerja akademik mahasiswa. Sehingga dapat memberikan lebih banyak perhatian kepada mahasiswa yang memiliki kinerja akademiknya buruk agar dapat lulus, berkualitas dan tepat waktu.

Beberapa penelitian untuk prediksi kinerja akademik sudah banyak dilakukan. Bhardwaj, B.K. and Pal, S. (2012) dan Aziz, A.A., dkk (2015) telah menerapkan NBC untuk prediksi kinerja akademik mahasiswa. NBC terbukti lebih unggul dari metode lain seperti penelitian yang dilakukan oleh Osmanbegovic E dan Suljic M pada tahun 2012 menggunakan sampel

yang terdiri dari 257 mahasiswa dengan 12 atribut, NBC terbukti memiliki akurasi tertinggi dan waktu komputasi yang lebih cepat dibandingkan dengan Multilayer Perceptron dan J48. R. Rathipriya dan T. A. Razak pada tahun 2014, dengan menggunakan data sebanyak 60 mahasiswa NBC terbukti memiliki akurasi tertinggi sebesar 74,57% dibandingkan dengan J48, ID3, IB1, dan OneR [9]. Pada tahun yang sama, Durairaj M. dan Vijitha C. membuktikan NBC memiliki akurasi tertinggi dibandingkan K-Means dan Decision Tree dengan 38 sampel yang terdiri dari 12 atribut. Mueen A, dkk pada tahun 2016 dengan 60 data mahasiswa membuktikan NBC memiliki akurasi tertinggi dengan atau tanpa algoritma seleksi atribut dibandingkan dengan Multilayer Perceptron dan C4.5/J48.

Namun Kabakchieva D pada tahun 2013 membuktikan NBC menghasilkan akurasi terendah dibandingkan dengan J48, JRip, K-NN, dan BayesNet dengan menggunakan data berjumlah 10330 mahasiswa dengan 20 atribut. Akurasi tertinggi dicapai oleh J48 dengan 66%. Sedangkan penelitian Dursun Delen tahun 2010, *Artificial Neural Network (MLP)*, *Decision Tree (CH)*, *Logistic regression* dikalahkan oleh *support vector machine (SVM)* dengan 81,18%. Penelitian ini menggunakan sebanyak 16.066 data siswa tahun 2004-2008 dari berbagai universitas di mid-west region U.S. Pada tahun 2013, Huang, S., & Fang, N dalam penelitiannya membuktikan bahwa SVM memiliki akurasi tertinggi mengalahkan *multiple linear regression (MLR)*, *multilayer perception (MLP) network*, dan *radial basis function (RBF) network*.

Dari beberapa penelitian di atas algoritma yang terbukti unggul dengan data yang berdimensi rendah adalah NBC. Untuk data yang berdimensi tinggi c4.5/J.48 memiliki akurasi tertinggi mengalahkan NBC. Sedangkan SVM terbukti dapat unggul pada data berdimensi tinggi. Namun pada penelitian (Chen, L. F., dkk, 2012) yang menggunakan 8 dataset yang

bervariasi SVM juga memperlihatkan hasil akurasi yang bervariasi. Terlihat jelas masalah klasifikasi umumnya ada pada jumlah atribut, tidak semua atribut sama pentingnya untuk tugas tertentu. Terlalu banyak atribut yang tidak penting akan meningkatkan waktu komputasi dan mengurangi akurasi. Agar proses akurasi lebih cepat dan akurat maka dibutuhkan strategi pemilihan atribut untuk mengetahui atribut yang tidak relevan terhadap kinerja klasifikasi (Chen, L. F., dkk, 2012).

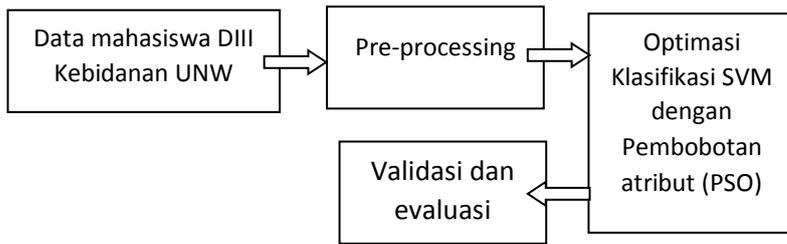
Algoritma pemilihan atribut diantaranya *Particle Swarm Optimization (PSO)* dan *Genetic Algorithm (GA)*. Ghamisi, P., & Benediktsson, J. A dan Chen, L. F., Su, C. T., Chen, K. H., & Wang, P. C (Chen, L. F., Su, C. T., membuktikan PSO yang diterapkan untuk seleksi atribut memiliki akurasi lebih tinggi dibandingkan dengan GA. Namun penelitian ini akan mencoba meningkatkan kinerja SVM dengan Algoritma Genetika (GA) untuk klasifikasi kinerja akademik mahasiswa agar dapat diketahui mahasiswa yang memiliki kinerja akademik baik dan buruk. Hasil penelitian dapat digunakan pihak pengelola institusi perguruan tinggi sebagai dasar untuk membuat kebijakan khusus bagi mahasiswa yang berkinerja buruk agar dapat tetap lulus tepat waktu, sehingga tidak mengurangi penilaian akreditasi perguruan tinggi.

## METODE PENELITIAN

### A. Pengumpulan Data

Penelitian ini akan menggunakan data mahasiswa DIII kebidanan Universitas Ngudi Waluyo yang lulus tahun 2014-2016. Data diambil dari dokumentasi program studi yang berupa file .xls. Data terdiri dari 3 file yang masing-masing berisi data pribadi mahasiswa, data nilai IPS dan IPK, dan data nilai tes osce. Atribut yang akan digunakan berjumlah 15 atribut dengan 14 atribut independen dan 1 atribut dependen. Total data berjumlah 322 *record*.

## B. KERANGKA KERJA



Gambar 3.1 Skema Kerja

### 1. Pre-processing

Dalam tahap ini akan dilakukan penggabungan data dari 3 file yang didapatkan berdasarkan nama mahasiswa. Hanya 15 atribut yang digunakan yaitu :

- a. Asal, merupakan atribut yang menjelaskan tentang daerah asal mahasiswa. Terdiri dari 3 kategori yaitu dalam kota, luar kota, dan luar propinsi.
- b. Status Sekolah, merupakan atribut yang menjelaskan status sekolah mahasiswa. Terdiri dari 2 kategori yaitu negeri dan swasta.
- c. Umur, merupakan atribut yang menjelaskan umur mahasiswa pada saat awal masuk kuliah. Umur merupakan atribut kontinu sehingga di kategorikan menjadi 2 sebagai berikut :
  - Umur 17-20 tahun masuk kategori standar.
  - Umur >20 tahun masuk kategori tua.

- d. IPS 1, berisi nilai indeks prestasi semester 1.
- e. IPS 2, berisi nilai indeks prestasi semester 2.
- f. IPS 3, berisi nilai indeks prestasi semester 3
- g. IPK, berisi nilai indeks prestasi kumulatif saat kelulusan. Atribut ini akan dijadikan atribut dependen.  $IPK < 3$  termasuk kategori kinerja buruk.  $IPK \geq 3$  termasuk kategori kinerja baik.
- h. Osce 1, berisi nilai tes osce tahap 1.
- i. Osce 2, berisi nilai tes osce tahap 2.
- j. Osce 3, berisi nilai tes osce tahap 3.
- k. Osce 4, berisi nilai tes osce tahap 4.
- l. Osce 5, berisi nilai tes osce tahap 5.
- m. Osce 6, berisi nilai tes osce tahap 6.
- n. Osce 7, berisi nilai tes osce tahap 7.
- o. Osce 8, berisi nilai tes osce tahap 8.

Jika terdapat *missing value* dan *noise* pada data akan diatasi dengan menghapus data tersebut secara manual. Proses ini menggunakan alat bantu Ms. Excel dan data yang sudah tergabung disimpan dengan nama “dataset” bertipe .xls.

### 2. Pemodelan menggunakan SVM-PSO

Implementasi SVM pada tahap ini menggunakan 4 kernel yang sering digunakan linier, polinomial, sigmoid dan radial.

Percobaan pertama menerapkan kernel-kernel tersebut. Setelah ditemukan kernel dengan akurasi tertinggi, kemudian percobaan kedua mengubah parameter C dengan nilai antara 0,1 sampai dengan 1 dan antara 5 sampai dengan 35. Hasil percobaan dengan merubah beberapa parameter tersebut dibandingkan dan diambil akurasi yang paling tinggi. Setelah itu GA diimplementasikan untuk meningkatkan akurasi SVM.

3. Validasi dan evaluasi

Dalam tahap ini dilakukan validasi dan pengukuran keakuratan hasil yang dicapai oleh model menggunakan teknik *confusion matrix* dan kurva ROC untuk pengukuran akurasi model.

HASIL PERCOBAAN

Untuk SVM dengan GA, dilakukan 4 percobaan yaitu membandingkan 3 kernel SVM, mengubah parameter C SVM, kemudian dengan SVM-GA.

Tabel 4.1 Hasil perbandingan akurasi percobaan pertama

Kernel	Akurasi	AUC
dot/linier	91.41	0.958
radial	85.74	0.896
polynomial	83.62	0.889
neural/sigmoid	86.25	0.886

Tabel 4.2 Hasil perbandingan akurasi percobaan kedua

C	Akurasi	AUC
0	91.41	0.958
0.1	90.53	0.965
0.2	90.96	0.957
0.3	90.96	0.957
0.4	90.98	0.974
0.5	91.38	0.974
0.6	91.38	0.968

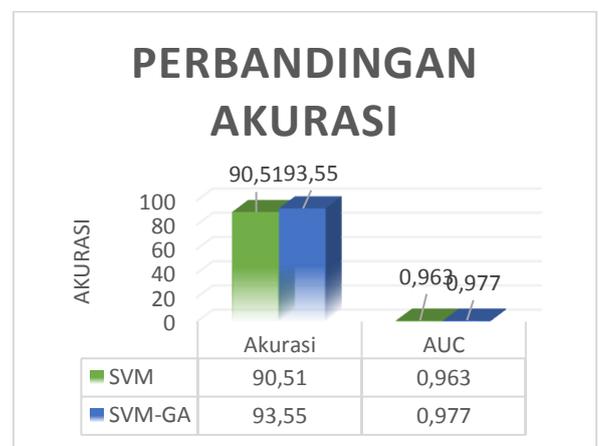
0.7	91.43	0.968
0.8	91.38	0.968
0.9	90.49	0.966
1	90.98	0.97

Tabel 4.3 Hasil perbandingan akurasi percobaan ketiga

Metode	Akurasi	AUC
SVM	90.51	0.963
SVM-GA	93.55	0.977

PEMBAHASAN

Hasil dari seluruh percobaan menghasilkan akurasi seperti tercantum pada grafik dibawah ini :



Gambar 4.1 Hasil Perbandingan akurasi

Terlihat bahwa metode yang diusulkan pada penelitian ini memiliki akurasi tertinggi dibandingkan dengan metode lainnya yaitu dengan akurasi 93,55% dan AUC 0,977. Metode SVM sebelumnya memiliki akurasi sebesar 90,51% dan AUC 0,963. Berdasarkan nilai AUC nya, performa dari metode SVM-GA yang diusulkan masuk dalam kelompok “Sempurna”.

PENUTUP

Penelitian ini berhasil menerapkan meningkatkan akurasi SVM dengan GA pada data kinerja akademik mahasiswa DIII Kebidanan Universitas Ngudi Waluyo angkatan 2012-2014. Hasil akurasi SVM-GA tertinggi yaitu dengan akurasi 93,55% dan AUC 0,977. Metode SVM sebelumnya memiliki akurasi sebesar 90,51% dan AUC 0,963. Berdasarkan nilai AUC nya, performa dari metode SVM-GA yang diusulkan masuk dalam kelompok “Sempurna”.

Dengan demikian untuk klasifikasi data kinerja akademik, metode SVM-GA yang diusulkan terbukti memiliki akurasi yang lebih tinggi dibandingkan dengan metode SVM jika digunakan pada data kinerja akademik mahasiswa DIII Kebidanan Universitas Ngudi Waluyo angkatan 2012-2014.

Untuk penelitian selanjutnya sebaiknya menambahkan jumlah data dan atribut yang mungkin berpengaruh pada kinerja akademik dari program studi lain. Dan mencoba membandingkan dengan metode klasifikasi dan metode optimasi lainnya.

#### REFERENSI

- Aziz, A.A., Ismail, N.H., Ahmad, F. and Hassan, H., (2015). A FRAMEWO FOR STUDENTS'ACADEMIC PERFORMANCE ANALYSIS USING NAÏVE BAYES CLASSIFIER. *Jurnal Teknologi Sciences & Engineering*), 75(3), pp.13-19.
- Bhardwaj, B.K. and Pal, S., (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*.
- Bramer, Max. (2007). *Principles of Data Mining*. London: Springer.
- Chen, L. F., Su, C. T., Chen, K. H., & Wang, P. C. (2012). Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis. *Neural Computing and Applications*, 21(8), 2087-2096.
- Durairaj, M. and Vijitha, C., (2014). Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), pp.5987-5991.
- Dursun Delen.(2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*. Volume 49. Issue 4. Pages 498-506. ISSN 0167-9236.
- Ghamisi, P., & Benediktsson, J. A. (2015). Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geoscience and Remote Sensing Letters*, 12(2), 309-313.
- Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer.
- Handayanna, F. (2012). Penerapan Particle Swarm Optimization Untuk Seleksi Atribut Pada Metode Support Vector Machine Untuk Prediksi Penyakit Diabetes. Tesis Magister Ilmu Komputer. Sekolah Tinggi Managemen Informatika dan Komputer Nusa Mandiri.

- Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, 61, 133-145.
- Kabakchieva, D., (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, 13(1), pp.61-72.
- Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36.
- Osmanbegovic, E. and Suljic, M., (2012). Data mining approach for predicting student performance. *Economic Review*, 10(1), pp.3-12.
- Peraturan BAN-PT No 4 Tahun 2017 tentang Instrumen Akreditasi.
- R. Rathipriya dan T. A. Razak, (2014). Predicting The Model For Academic Performance Using Classification Techniques. *International Journal of Computer Engineering and Applications*, 8(1).
- Sahu, B., & Mishra, D. (2012). A novel feature selection algorithm using particle swarm optimization for cancer microarray data. *Procedia Engineering*, 38, 27-31.
- Vercellis, Carlo (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Wiley & Sons, Ltd.
- Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And Techniques*. Burlington, Usa: Morgan Kaufmann Publishers.
- Zeniarta, J. (2012). Opinion Mining of Movie Review on Twitter Using Support Vector Machine