

Klasifikasi Prestasi Akademik Mahasiswa Berdasar Hasil Tes Potensi Akademik Menggunakan *Support Vector Machine*

Agustinus Handaya Ajitama^{a,*}, Paulina H. Prima Rosa^{b,*}

^{a,b} Prodi Informatika Universitas Sanata Dharma, Yogyakarta

*ajitomo003@gmail.com , rosa@usd.ac.id

Abstract

This paper describes research that aims to classify student academic achievement based on the results of the Academic Potential Test (APT), using data containing APT scores and Grade Point Average (GPA) in semester 1 to semester 4. To get the best accuracy, some experiments were carried out by varying the fold, kernel, and multiclass. The folds used in the experiment are 5, 7, and 9 folds. The kernel used is a linear kernel, Gaussian Radial Basic Function (RBF), and polynomial. Meanwhile, the multiclass used in this experiment is one against one and one against all. The highest accuracy of 80% was obtained in the GPA classification in semester 4 using multiclass one against one and one against all, RBF kernel, and 9-fold cross validation

Keywords: *Data Mining*; Classification, Support Vector Machine, Academic Potential Test

Abstrak

Makalah ini menjabarkan tentang penelitian yang bertujuan untuk mengklasifikasi prestasi akademik mahasiswa berdasarkan hasil Tes Potensi Akademik (TPA), menggunakan data berisi nilai TPA dan Indeks Prestasi Kumulatif (IPK) semester 1 hingga IPK semester 4. Untuk mendapatkan akurasi terbaik, eksperimen dilakukan dengan memvariasikan *fold*, *kernel*, dan *multiclass*. Fold yang digunakan dalam eksperimen ini yaitu 5, 7, dan 9 *fold*. *Kernel* yang digunakan adalah kernel *linear*, *Gaussian Radial Basic Function* (RBF), dan *polynomial*. Sementara *multiclass* yang digunakan dalam eksperimen ini yaitu *one against one* dan *one against all*. Akurasi tertinggi sebesar 80% diperoleh pada klasifikasi IPK di semester 4 dengan menggunakan *multiclass one against one* maupun *one against all*, *kernel RBF*, dan *9-fold cross validation*.

Kata Kunci (Arial, 9, Bold): *Data Mining*, Klasifikasi, *Support Vector Machine*, Tes Potensi Akademik.

1. Pendahuluan

Pendidikan tinggi umumnya diawali dengan seleksi masuk untuk memastikan bahwa mahasiswa tidak mengalami kesulitan selama proses studinya. Salah satu alat bantu untuk seleksi adalah melalui Tes Potensi Akademik (TPA), sebagaimana yang dipergunakan oleh Universitas Sanata Dharma. Hasil TPA tersebut juga dapat dipergunakan untuk mengklasifikasi dan memprediksi prestasi akademik mahasiswa yang bermanfaat untuk pendampingan akademik mahasiswa yang bersangkutan selama proses perkuliahan.

Dalam makalah ini, penulis mengimplementasikan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan prestasi akademik mahasiswa. Dilakukan beberapa eksperimen untuk mengidentifikasi akurasi terbaik dari metode SVM dalam mengklasifikasi data prestasi mahasiswa program studi Pendidikan

Fisika. Beberapa penelitian sebelumnya beberapa peneliti pernah melakukan klasifikasi prestasi akademik seperti Farida & Ratih (2017) yang mengevaluasi Prestasi Akademik Mahasiswa Universitas Nusantara PGRI Kediri dengan Algoritma Naïve Bayes, menghasilkan akurasi sebesar 66,6%. Penelitian lainnya dilakukan oleh Sabna & Muhandi (2015) yang memprediksi prestasi akademik mahasiswa berdasarkan dosen, motivasi, kedisiplinan, ekonomi, dan hasil belajar. Kurniawan (2019) juga sudah melakukan penelitian untuk memprediksi prestasi akademik mahasiswa dengan menggunakan Algoritma *K-Nearest Neighbour*, dengan akurasi 70%.

2. Kajian Literatur

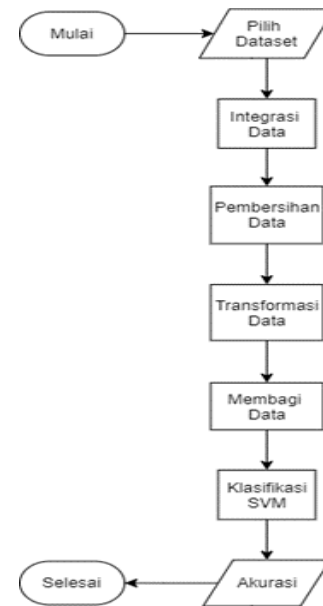
Klasifikasi merupakan proses untuk menemukan sebuah model atau fungsi yang dapat mendeskripsikan dan membedakan

kelas data atau konsep (Jiawei, et.al., 2012). Algoritma *Support Vector Machine* (SVM) merupakan salah satu algoritma klasifikasi yang diperkenalkan oleh Vapnik pada tahun 1992 yang mula-mula dipergunakan untuk klasifikasi 2 kelas. Dalam perkembangannya, dikembangkan SVM untuk *multiclass*. Terdapat dua pendekatan dalam mengimplementasikan SVM *multiclass*. Pendekatan pertama, menggabungkan berapa SVM biner. Pendekatan kedua, menggabungkan semua data dari semua kelas ke dalam sebuah bentuk permasalahan optimasi (Suyanto,2019). Beberapa contoh metode yang menerapkan pendekatan pertama adalah *one-against-all* dan *one-against-one* (Suyanto,2019).

Metode *Support Vector Machine* pernah dipergunakan dalam beberapa penelitian dan menunjukkan akurasi yang baik. Salah satunya adalah Windrawati (2020) yang meneliti klasifikasi varietas kopi Arabika dan menghasilkan akurasi sebesar 48,33% dengan menggunakan kernel polynomial. Penelitian lainnya dilakukan Octaviani (2014) untuk klasifikasi data akreditasi Sekolah Dasar (SD) di Kabupaten Magelang yang menghasilkan akurasi 100% ketika menggunakan fungsi kernel RBF dan 98,810% ketika menggunakan fungsi kernel Polynomial. Hasibuan (2017) meneliti klasifikasi diagnosa penyakit Demam Berdarah Dengue (DBD) dan mendapatkan hasil akurasi sebesar 96,4286% dengan menggunakan kernel polynomial. Kurniawaty (2018) melakukan penelitian klasifikasi gangguan jiwa skizofrenia dan menghasilkan akurasi sebesar 100% dengan kernel polynomial. Dalam penelitian ini akan diidentifikasi apakah akurasi SVM juga baik jika diterapkan pada kasus klasifikasi prestasi akademik mahasiswa.

3. Metode Penelitian

Tahap-tahap penelitian digambarkan dalam *flowchart* dalam gambar 1. Dalam penelitian ini digunakan *dataset* program studi Pendidikan Fisika Universitas Sanata Dharma. Data nilai Tes Potensi Akademik terdiri dari penalaran verbal, kemampuan numerik, penalaran mekanik, hubungan ruang, Bahasa Inggris, dan Indeks Prestasi Kumulatif (IPK) dari semester satu sampai empat. Tabel 1 berisi penjelasan setiap atribut nilai yang akan menjadi isi atribut pada file tersebut.



Gambar 1. Skema Penelitian

Tabel 1. Penjelasan Atribut dalam Dataset

No	Nama Atribut	Penjelasan	Nilai
1	Penalaran Verbal	Nilai Tes Potensi Akademik bagian Penalaran Verbal	1 - 10
2	Kemampuan Numerik	Nilai Tes Potensi Akademik bagian Kemampuan Numerik	1 - 10
3	Penalaran Mekanik	Nilai Tes Potensi Akademik bagian Penalaran Mekanik	1 - 10
4	Hubungan Ruang	Nilai Tes Potensi Akademik bagian Hubungan Ruang	1 - 10
5	Bahasa Inggris	Nilai Tes Potensi Akademik bagian Bahasa Inggris	1 - 10
6	IPK 1	IPK Semester 1	0 - 4
7	IPK 2	IPK Semester 2	0 - 4
8	IPK 3	IPK Semester 3	0 - 4
9	IPK 4	IPK Semester 4	0 - 4

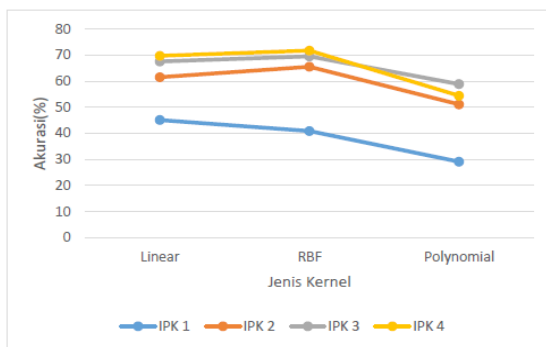
Pada tahap *preprocessing* dilakukan proses pembersihan data dengan metode deteksi *outlier*, integrasi data, transformasi data. Pembagian data untuk validasi dilakukan dengan metode *cross-validation*. Dalam tahap klasifikasi diterapkan algoritma SVM dengan memanfaatkan *library* pada *Python* menggunakan fungsi biner yang ditambahkan fungsi *multiclass*. Diterapkan dua metode *multiclass* yaitu metode *one-against-one* dan metode *one-against-all*.

Dalam penelitian dilakukan beberapa macam eksperimen dengan melakukan variasi pada *fold*, *kernel*, dan *multiclass*. *Fold* yang digunakan adalah 5, 7, dan 9. *Kernel* yang digunakan yaitu *linear*, *RBF*, dan *polynomial*. *Multiclass* yang digunakan yaitu *one against one* dan *one against all*.

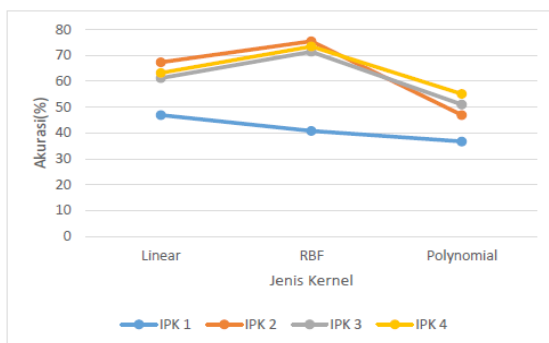
Algoritma SVM diimplementasikan dengan menggunakan Python.

4. Hasil dan Pembahasan

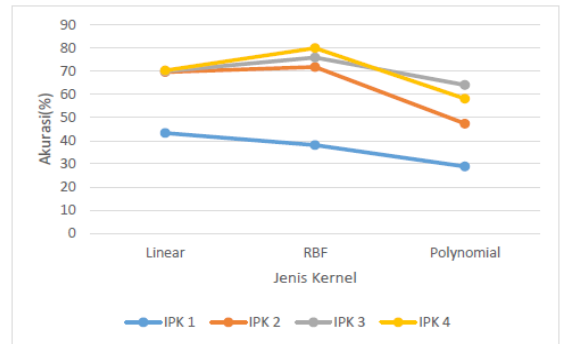
Gambar 2 hingga gambar 4 berikut ini menunjukkan grafik hasil eksperimen dengan menerapkan *one against one* pada dataset dengan jumlah fold 5, 7, dan 9.



Gambar 2. Grafik akurasi menggunakan 5-Fold, *Multiclass One-Against-One*



Gambar 3. Grafik akurasi menggunakan 7-Fold, *Multiclass One-Against-One*

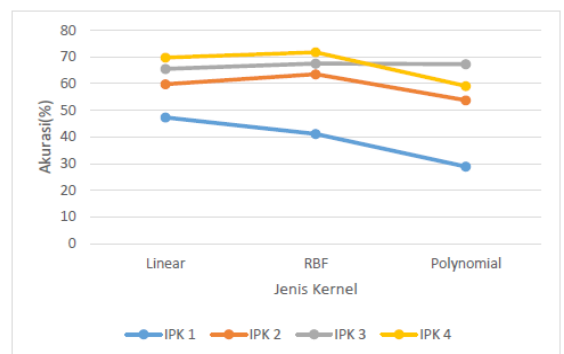


Gambar 4. Grafik akurasi menggunakan 9-Fold, *Multiclass One-Against-One*

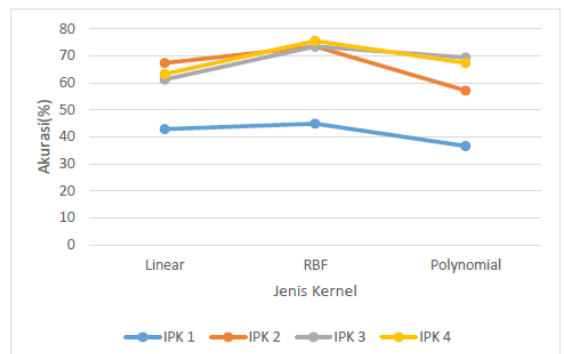
Dengan menggunakan *Multiclass One-Against-One* dalam gambar 2 hingga gambar 4, dalam setiap fold, terlihat bahwa klasifikasi pada IPK semester 1 selalu paling rendah akurasinya dibandingkan klasifikasi pada IPK semester 2, 3, dan 4.

Gambar di atas juga menunjukkan bahwa pada IPK semester 2,3, dan 4, kernel RBF tampak lebih baik akurasinya dibandingkan kernel lainnya, namun hal itu tidak terjadi pada IPK semester 1.

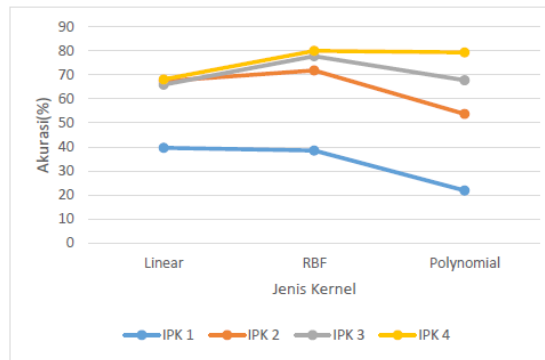
Gambar 5 hingga gambar 8 berikut ini menunjukkan grafik hasil eksperimen dengan menerapkan *one against all* pada dataset dengan jumlah fold 5, 7, dan 9.



Gambar 5. Grafik akurasi menggunakan 5-Fold, *Multiclass One-Against-All*



Gambar 6. Grafik akurasi menggunakan 7-Fold, *Multiclass One-Against-All*



Gambar 7. Grafik akurasi menggunakan 9-Fold, Multiclass One-Against-All

Dari gambar 5 hingga 7, serupa dengan Multiclass One-Against-One dalam Multiclass One-Against-All terlibat pola akurasi klasifikasi IPK semester 1 selalu di bawah akurasi klasifikasi IPK semester lainnya. Kernel RBF juga memiliki performansi yang cenderung lebih baik dibandingkan kernel lainnya, kecuali pada IPK semester 1.

Dari hasil eksperimen di atas, untuk setiap atribut IPK yang dijadikan label kelas, dapat diidentifikasi akurasi tertingginya seperti dalam tabel 2 berikut ini

Tabel 2. Akurasi Terbaik pada Tiap Atribut IPK

IPK	Akurasi	Multiclass	Kernel	k-fold
1	47,333%	one against all	linear	5
2	75,51%	one against one	RBF	7
3	77,778%	one against all	RBF	9
4	80%	one against one dan one against all	RBF	9

Secara keseluruhan, tampak bahwa akurasi terbaik diperoleh pada klasifikasi IPK semester 4 sebesar 80% pada kedua jenis multiclass menggunakan kernel RBF pada 9-fold cross validation. Rendahnya akurasi klasifikasi pada semester 1 menimbulkan beberapa kemungkinan dugaan, misalnya semester 1 masih merupakan masa adaptasi bagi mahasiswa sehingga belum menunjukkan prestasi yang sesuai dengan potensi akademiknya. Semakin tinggi semesternya, semakin baik akurasi klasifikasinya. Ada kemungkinan mahasiswa sudah berhasil

beradaptasi seiring dengan bertambahnya semester. Pendalaman lebih lanjut terkait dugaan ini perlu dilakukan melalui penelitian berikutnya.

5. Kesimpulan dan Saran

Dapat disimpulkan bahwa algoritma Support Vector Machine dapat diimplementasikan untuk klasifikasi prestasi akademik mahasiswa prodi Pendidikan Fisika USD berdasarkan hasil tes potensi akademik. Akurasi klasifikasi terbaik yang dihasilkan oleh metode Support Vector Machine terhadap dataset prodi Pendidikan Fisika adalah 80% pada IPK 4 dengan multiclass one against one dan one against all, kernel RBF, dan 9-fold cross validation. Dengan demikian, klasifikasi ini dapat dimanfaatkan oleh pengelola prodi untuk meningkatkan pendampingan mahasiswa dalam masa studinya. Penelitian lebih lanjut dapat dilakukan untuk memahami temuan hasil klasifikasi ini.

Daftar Pustaka

Farida, Intan Nur & Ratih Kumalasari N., (2017), Penggunaan Algoritma Naive Bayes Untuk Mengevaluasi Prestasi Akademik Mahasiswa Universitas Nusantara PGRI Kediri. Kediri: Universitas Nusantara PGRI Kediri, Jurnal Sains dan Informatika Volume 3, Nomor 2, November 2017, e-ISSN: 2598-5841.

Han, J., Kamber, M. & Pei, J., (2012), Data Mining: Concepts and Techniques. Waltham: Morgan Kaufmann Publishers.

Hasibuan, Chainur A. et.al., (2017), Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) Menggunakan Support Vector Machine (SVM) Berbasis GUI Matlab. Jurnal Gaussian, Volume 6, Nomor 2, Tahun 2017, Halaman 171-180, ISSN: 2339-2541.

Kurniawan, Vincentius B., (2019), Prediksi Prestasi Akademik Mahasiswa Berdasarkan Hasil Tes Potensi Akademik dengan Algoritma K-Nearest Neighbor, Skripsi. Yogyakarta: Universitas Sanata Dharma

Kurniawaty, D. et.al., (2018), Klasifikasi Gangguan Jiwa Skizofrenia Menggunakan Algoritme Support Vector Machine (SVM). Universitas Brawijaya, Jurnal

Pengembangan Teknologi Informasi dan Ilmu Komputer Vol.2, No.5, Mei 2018, halaman 1866-1873, e-ISSN: 2548- 964X.

Ocataviani, Pusphita A. et.al., (2014), Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang. Jurnal Gaussian, Volume 3, Nomor 4, Tahun 2014, Halaman 811-820, ISSN: 2339-2541.FLEXChip Signal Processor (MC68175/D), Motorola, 1996

Suyanto, (2019), Data Mining untuk Klasifikasi dan Klasterisasi Data. Bandung: penerbit INFORMATIKA.