

KOMPARASI KERNEL PADA ALGORITMA SUPPORT VECTOR MACHINE STUDI KASUS KLASIFIKASI PENJURUSAN DI SMA SAVERIUS SRAGEN

Theopilus Bayu S¹⁾, Adhistrya Erna Permanasari²⁾, Indriana Hidayah³⁾

Teknik Elektro dan Teknologi Informasi UGM Yogyakarta

Jl. Grafika No 2, Yogyakarta

Email : theopilusbayu.mti.18a@mail.ugm.ac.id¹⁾, adhistrya@ugm.ac.id²⁾, ana_phy@te.ugm.ac.id³⁾

Abstrak

Support Vector Machine digunakan untuk mengklasifikasikan penjurusan pada jenjang SMA. Dataset terdiri dari 40 siswa-siswi. Setiap siswa-siswi memiliki 15 atribut. Dengan menggunakan metode pendekatan datamining, yaitu metode klasifikasi Support Vector Machine dan perangkat lunak Rapidminer, akan dianalisa efek dari perubahan fungsi kernel variasi faktor pinali (C) terhadap akurasi model klasifikasi. Akurasi terbaik adalah pada kernel linear dengan parameter pinalti (C) 0.1 sebesar 88.89%.

Kata kunci — Algoritma Support Vector Machine, klasifikasi penjurusan, fungsi kernel.

1. PENDAHULUAN

Membimbing siswa-siswi untuk memilih jurusan yang tepat sangat penting dalam jenis pembelajaran apapun. Baik itu pembelajaran konvensional maupun pembelajaran jarak jauh. Keterlambatan mengetahui jurusan yang tepat bagi siswa-siswi adalah suatu kerugian, baik siswa maupun bagi lembaga yang bersangkutan.

Sistem untuk melakukan pengklasifikasian jurusan saat ini pada studi kasus di SMA Saverius Sragen masih menggunakan pengklasifikasian secara manual, artinya sekolah menentukan pengklasifikasian jurusan pada siswa-siswi kelas X dengan mengumpulkan nilai *raport* semester 1 dari masing-masing wali kelas X dan hasil nilai tes psikologi. Setelah data dikumpulkan guru bimbingan konseling menganalisa dan menghitung hasilnya secara manual serta memcokkannya dengan kriteria jurusan tertentu. Hal ini menyulitkan guru bimbingan konseling dalam melakukan pengklasifikasian penjurusan serta membutuhkan waktu yang lama untuk menganalisa dan menghasilkan keputusan.

Sudah ada alat bantu untuk melakukan prediksi maupun klasifikasi dengan kasus di lembaga pendidikan salah satunya adalah menggunakan metode *Support Vector Machine* [1]. SVM sebagai salah satu metode klasifikasi yang memiliki tingkat akurasi yang tinggi dalam melakukan prediksi pengklasifikasian potensi pada lembaga pendidikan.

SVM memiliki kelebihan dalam mengklasifikasikan suatu pola dengan akurat walaupun dengan keterbatasan *dataset* seperti pada penelitian SVM yang diterapkan dalam lingkup kasus lembaga pendidikan yang dilakukan oleh Nimit Pattanasri, dkk pada tahun 2012 [2].

Penelitian yang bertujuan mengklasifikasikan jenis *slide* presentasi berdasarkan fitur pada tiap matakuliah ini hanya menggunakan data sebanyak 102 kuisisioner yang diberikan kepada mahasiswa.

SVM memiliki Kemampuan memecahkan masalah berdimensi tinggi walaupun dengan keterbatasan sampel [3]. Hal ini mereka buktikan melalui penelitian dengan mengklasifikasikan spesialisasi mahasiswa yang cocok melalui ekstrasi pola dari sinyal *electroencephalography* (EEG). Pernyataan ini juga didukung oleh Vapnik. Vapnik adalah salah satu pencetus dari SVM. Vapnik mengungkapkan bahwa tingkat memecahkan masalah berdimensi tinggi pada SVM tidak terlalu dipengaruhi oleh dimensi dari *input vector* [4]. Vapnik juga mengungkapkan bahwa SVM memiliki performansi yang bagus walaupun dengan data yang bersifat *non linear* [4]. Walaupun SVM memiliki berbagai kelebihan, SVM juga memiliki kelemahan diantaranya adalah pemilihan kernel yang tepat sesuai dengan studi kasus masih dalam penelitian lebih lanjut [5], perlunya dilakukan penelitian untuk metode penentuan parameter SVM seperti pernyataan Jacobus dengan membuat Sistem Deteksi Intrusi Jaringan dengan SVM dilihat bahwa penentuan parameter SVM yang tepat dapat memberikan model dengan performansi yang baik [6].

1.1 METODOLOGI

A. Alat dan Bahan

Bahan material yang digunakan pada *paper* ini meliputi *dataset* nilai *raport* semester 1 siswa-siswi kelas X SMA Saverius tahun ajaran 2010/2011 yang meliputi 40 orang siswa-siswi, beserta nilai psikotest penjurusan masing-masing siswa. Alat yang digunakan adalah laptop, dan perangkat lunak Rapid miner *community edition* yang digunakan untuk menerapkan metode SVM.

B. Algoritma SVM

Pada pembahasan ini akan dibahas mengenai algoritma yang dipakai pada penelitian ini yaitu SVM. Jika terdapat i data pada dua kelas yang terpisah secara *linear*, maka label tiap data di formuliskan sebagai $y_i \in \{-1, 1\}$. Klasifikasi terhadap data dapat diformulasikan dengan persamaan 1:

$$w \cdot x_i + b \geq 1 \text{ untuk } y_i = +1 \quad (1)$$
$$\text{dan } w \cdot x_i + b \leq -1 \text{ untuk } y_i = -1.$$

Tujuan dari *Support vector machine* adalah mencari jarak y terbesar dari persamaan 1. Pencarian jarak terbesar ini dapat menggunakan teknik *Lagrange*

Multiplier yang akan menghasilkan fungsi keputusan SVM menjadi:

$$f(x_d) = \sum_{i=1}^{n_s} \alpha_i y_i x_i x_d + b \quad (2)$$

Namun pada kenyataannya data yang sering kita jumpai bersifat *non linear*. Pada kasus seperti ini kedua bidang pembatas seperti pada persamaan 1 harus diubah sehingga lebih fleksibel dengan melakukan penambahan variabel *slack* (ξ) menjadi seperti persamaan 3:

$$w \cdot x_i + b \geq 1 - \xi \text{ untuk } y_i = +1 \quad (3)$$

$$\text{dan } w \cdot x_i + b \leq -1 + \xi \text{ untuk } y_i = -1.$$

Pencarian bidang pemisah terbaik dengan variabel *slack* (ξ) disebut *soft margin hyperplane*. Dengan demikian formulasi pencarian bidang pemisah terbaik menjadi:

$$\min \frac{1}{2} |w|^2 + C \sum_{i=1}^n \xi_i \quad (4)$$

C adalah parameter yang menentukan besar dari pinalti akibat kesalahan dalam klasifikasi, Nilai C ditentukan oleh pengguna. Metode lain yang dapat digunakan untuk mengklasifikasikan data yang bersifat *non linear* adalah dengan mentransformasikan data tersebut ke dalam dimensi ruang fitur. Caranya, data dipetakan dengan menggunakan fungsi pemetaan (transformasi) $x_k \rightarrow \phi(x_k)$ ke dalam ruang fitur. Melalui fungsi transformasi ini, hasil pembelajaran yang dihasilkan adalah :

$$f(x_d) = \sum_{i=1}^{n_s} \alpha_i y_i \phi(x_i) \phi(x_d) + b \quad (5)$$

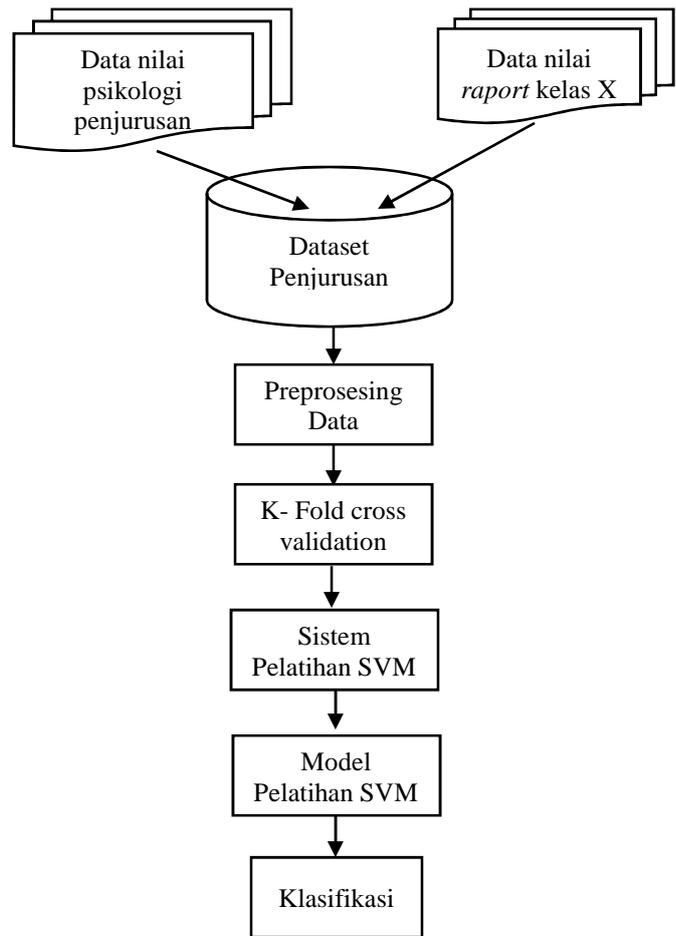
Namun penggunaan ruang fitur mengakibatkan komputasi yang sangat besar. Untuk mengatasi masalah ini SVM menggunakan *kernel trick* ($\phi(x_i) \phi(x_d)$). Fungsi yang dihasilkan dari pelatihan adalah :

$$f(x_d) = \sum_{i=1}^{n_s} \alpha_i y_i K(x_i, x_d) + b \quad (6)$$

Pada paper ini digunakan lima kernel untuk dilakukan perbandingan akurasi yaitu dot kernel, radial, polynomial, neural, epacheninikov, Gaussian_combination, dan multiquadratic kernel.

C. Pembangunan Model SVM

Pembangunan model menggunakan model klasifikasi SVM. Data masukan terdiri dari dua masukan yaitu data nilai *raport* kelas X dan nilai psikologi siswa. Data nilai *raport* meliputi nilai mata pelajaran biologi, ekonomi, fisika, geografi, kimia, matematika, sejarah, sosiologi. Sedangkan data nilai psikologi meliputi nilai IQ, kecerdasan rasional, abstrak konseptual, analisa sintesa, logika verbal, dan logika numerik. Sedangkan keluaran yang dihasilkan adalah penjurusan SMA yaitu IPA dan IPS. Gambaran umum pembangunan model SVM pada gambar 1.0



Gambar 1 Pembangunan Model SVM

Dari gambar 1 dapat dilihat bahwa data nilai tes psikologi bakat penjurusan, data nilai *raport* kelas digabungkan menjadi *dataset* penjurusan yang kemudian dilakukan pre-prosesing data, salah satunya dengan melakukan *scaling method* dengan menggunakan *Min-Max normalization* hingga data siap digunakan. Penggunaan *min-max normalization* ini didasarkan oleh penelitian yang menunjukkan bahwa metode *min-max normalization* menunjukkan hasil terbaik dibandingkan dengan *scaling method* yang lain [6]. Data yang telah dilakukan pre-prosesing kemudian dibagi menjadi dua yaitu data *training* dan data *testing* dengan menggunakan teknik *k-fold cross validation*. *Dataset* bakat penjurusan yang telah dilakukan *k-fold cross validation* kemudian dilakukan proses *training* pada Sistem pelatihan SVM untuk menghasilkan model pelatihan yang cocok.

2. PEMBAHASAN

Pada paper ini untuk melihat performansi dari model yang dihasilkan digunakan 3 fold cross-validation. Menurut penelitian Ning 3 fold cross-validation cocok digunakan untuk data yang relatif kecil [7]. digunakan dataset sebanyak 40 siswa. Masing-masing siswa memiliki 15 atribut masukan yang digunakan untukl klasifikasi jurusan. Tujuan dari penelitian ini adalah menganalisa metode terutama pada pengaruh perubahan

fungsi kernel serta variasi faktor pinaliti (C) terhadap akurasi pengklasifikasian.

Pada tabel 1 memperlihatkan tingkat akurasi pengklasifikasian data menggunakan SVM.

Tabel 1 Tingkat akurasi klasifikasi SVM

Par C	Tipe Kernel				
	Dot	Radial	Poli- nomial	Neu- ral	Anova
0.1	88.89	58.33	63.89	80.56	80.56
0.2	86.11	58.33	63.89	75.00	80.56
0.4	77.78	58.33	63.89	80.56	80.56
0.8	77.78	58.33	63.89	69.44	80.56
1.0	77.78	58.33	63.89	75.00	80.56
1.2	77.78	58.33	63.89	72.22	80.56
1.4	77.78	58.33	63.89	72.22	80.56
1.8	77.78	58.33	63.89	66.67	80.56
Par C	Epac henini kov	Gaus siancomb ination	Multi quadr atic		
0.1	50.00	33.33	50.00		
0.2	50.00	33.33	50.00		
0.4	50.00	33.33	50.00		
0.8	50.00	33.33	50.00		
1.0	50.00	33.33	50.00		
1.2	50.00	33.33	50.00		
1.4	50.00	33.33	50.00		
1.8	50.00	33.33	50.00		

Pada tabel 1 memperlihatkan bahwa memperbesar nilai parameter C tidak memperbaiki tingkat akurasi yang diberikan. Hal ini sesuai dengan penelitian yang dilakukan oleh Jacobus [6]. Tiap kernel memiliki tingkat akurasi yang rendah, hal ini disebabkan oleh keterbatasan data pelatihan yang digunakan. Tingkat akurasi terbaik adalah 88.89% diperoleh dari tipe kernel dot dengan nilai parameter C sebesar 0.1

3. KESIMPULAN

Pada paper ini dipaparkan mengenai pengklasifikasian penjurusan pada siswa SMA dengan menggunakan pendekatan Support Vector Machine. Digunakan dataset relatif sangat sedikit, yaitu 40 siswa. Tiap instances siswa memiliki 15 atribut. penelitian ini adalah menganalisa metode terutama pada pengaruh perubahan fungsi kernel serta variasi faktor pinaliti (C) terhadap akurasi pengklasifikasian.

Hasil analisa menunjukkan bahwa Tingkat akurasi terbaik adalah 88.89% diperoleh dari tipe dot kernel dengan nilai parameter C sebesar 0.1.

Pengembangan penelitian yang akan datang perlu dilakukan dengan *dataset* yang relatif besar agar tingkat keakuratan performansi model terjamin. Perlu dilakukan mengenai metode pemilihan atribut, dan penentuan parameter SVM yang cocok.

DAFTAR PUSTAKA

- [1] F. Jui-Hsi, C. Jui-Hung, H. Yueh-Min, and C. Han-Chieh, *A Support Vector Regression based Prediction of Student's School Performance*, Computer, Consumer and Control (IS3C), International Symposium, pp.84-87, 2012.
- [2] P. Nimit, M. Masayuki, M. Michihiko, *Learning to Estimate Slide Comprehension in Classroom with Support Vector Machines*. IEEE Computer Society, vol. 5, pp.52-61, March. 2012.
- [3] Y. Zhang, and W. Wang, *Pattern Classification of Electroencephalography from the Typical Specialized Students*, Education Technology and Computer Science (ETCS), International Workshop., vol. 1, pp. 836-839, March. 2010.
- [4] C.Cortes, and V.Vapnik, *Support- VectorNetwork*, *Machine Learning*, Vol 20, No 3, pp 273 – 297, 1995.
- [5] C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Data mining and Knowledge Discovery, Vol 2.1998.
- [6] J. Agustinus, *Sistem Deteksi Intrusi Jaringan dengan metode Support Vector Machine*, Tesis, Jurusan Ilmu-Ilmu Komputer FMIPA UGM, Yogyakarta, 2012.
- [7] R. Ning, *Klasifikasi Siswa sebagai Prediksi Prestasi Akademik dengan menggunakan Neuro Fuzzy*, Skripsi, Jurusan Teknik Elektro dan Teknologi Informasi, UGM, Yogyakarta, 2013.

BIODATA PENULIS

Theopilus Bayu Sasongko, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknologi Informasi UKDW Yogyakarta, lulus tahun 2012. Kemudian langsung melanjutkan studi Strata dua hingga saat ini di jurusan Teknik Elektro & Teknologi Informasi di Universitas Gadjah Mada Yogyakarta.

Adhitya Erna Permanasari, memperoleh gelar Doctor di Universitas Teknologi Petronas (UTP) Malaysia. Saat ini menjadi Dosen di Universitas Gadjah Mada Yogyakarta, beliau memiliki konsentrasi minat penelitian yang berkaitan dengan Teknologi Basis data, dan penambangan data (*datamining*).

Indriana Hidayah, memperoleh gelar Master di Universitas Gadjah Mada (UGM) Yogyakarta. Saat ini menjadi Dosen di Universitas Gadjah Mada Yogyakarta, beliau memiliki konsentrasi minat penelitian yang berkaitan dengan dengan penambangan data (*datamining*).

