



## Penerapan Support Vector Machine Untuk Jalur Peminatan Program Studi

Dion Krisnadi<sup>1</sup>, Nadya D. Bachtiar<sup>2</sup>, Samuel Lukas<sup>3</sup>, Pujiyanto Yugopuspito<sup>4</sup>, Petrus Widjaja<sup>5</sup>  
<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan  
<sup>5</sup>Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Pelita Harapan  
Samuel.lukas@uph.edu

### Abstract

*It is common that in a curriculum of field of study provides some electives subjects to be selected by their students. Some electives subjects are grouped into a major study. Universitas Pelita Harapan Informatics department offers three major studies. They are Software Engineering, Medical Informatics and Intelligent & Interactive Media. These majors are provided in order to fit interest of the students. Students must choose one major for pursuing their degree after they finished the first two years studies. This paper discusses on how to implement the Support Vector machines technology for helping students in determining their major based on their academic records. The inputs of the system are grades of the subjects of the first two years studied and the outputs is the GPA of the electives subjects for each major.*

Keywords: *Major study, Informatics department, Support Vector Machines, Curriculum*

### Abstrak

Sudah umum bahwa pada kurikulum bidang studi tertentu, menyediakan beberapa mata kuliah pilihan untuk dipilih mahasiswa. Beberapa matakuliah pilihan dikelompokkan ke dalam peminatan tertentu. Di program studi Teknik Informatika Universitas Pelita Harapan menawarkan tiga Peminatan, peminatan Rekayasa Perangkat Lunak, Informatika Medis dan Media Cerdas & Interaktif. Peminatan ini disediakan untuk memenuhi minat para mahasiswa. Mahasiswa harus memilih satu peminatan untuk mendapatkan gelar mereka setelah mereka menyelesaikan pembelajaran di dua tahun pertama. Makalah ini membahas tentang bagaimana menerapkan teknologi *Vector Support Machine* untuk membantu siswa dalam menentukan peminatan mereka berdasarkan data akademis. Input sistem adalah prestasi belajar mahasiswa dari dua tahun pertama dan keluarannya adalah IPK dari mata pelajaran pilihan untuk masing-masing peminatan.

Kata kunci: Peminatan pembelajaran, Teknik Informatika, Support Vector Machine, Kurikulum.

### 1. Pendahuluan

Jalur peminatan memungkinkan mahasiswa dapat mengembangkan kemampuan akademis pada bidang studi yang ia pelajari sesuai dengan minat mereka masing-masing. Mahasiswa dapat memilih peminatan biasanya pada tahun ketiga setelah lulus sejumlah matakuliah di tahun pertama dan kedua. Sebagian mahasiswa memilih jalur peminatannya didasari atas minat mereka sendiri tanpa memperdulikan kemampuan akademisnya. Sebagian lagi mengacu kepada kemampuan akademisnya tanpa memperdulikan minatnya. Sudah tentu yang terbaik adalah memperhatikan kedua faktor yaitu minat dan kemampuan akademis secara proporsional. Lukas[1] sudah mengusulkan model penentuan jalur peminatan menggunakan fuzzy system. Makalah ini dibahas penerapan teknologi *Support Vector Machine*, SVM untuk menentukan jalur peminatan mahasiswa. Sistem akan menerima sejumlah prestasi belajar mahasiswa dari matakuliah di tahun pertama dan kedua untuk matakuliah program studinya, kemudian system akan

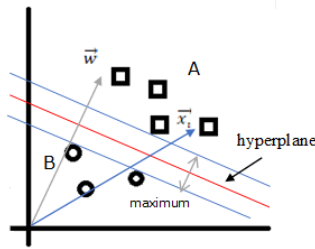
menghitung kemungkinan nilai Indeks Prestasi Kumulatif, IPK siswa apabila ia mengambil suatu jalur peminatan tertentu.

### 2. Teknologi Support Vector Machine

*Support Vector Machine* (SVM) dipakai untuk klasifikasikan atau regresi data pada *supervised machine learning*. Hasil klasifikasi dinyatakan dengan *hyperplane* [2]. Ilustrasi dari hasil pengklasifikasian data A dan data B yang diperlihatkan pada Gambar 1 dinyatakan dalam bentuk garis merah sebagai *hyperplane*. Garis merah itu memisahkan data ke kategori A dan B sedemikian rupa sehingga jarak terdekat titik data ke *hyperplane* maksimum.

Vector  $\bar{w}$  tegak lurus dengan bidang *hyperplane* memenuhi persamaan (1) dengan  $c_0$  adalah jarak dari titik origin ke bidang *hyperplane*. Karena jarak terdekat antara data A dan data B ke *hyperplane* dinyatakan dengan persamaan (2) maka untuk memaksimalkannya dapat dilakukan dengan meminimalkan  $\|w\|$  dengan batasan persamaan (3) dimana  $b_i = 1$  untuk  $\forall \bar{x}_i \in A$  dan

$b_i = 0$  untuk  $\forall \bar{x}_i \in B$ . Multiple langrange digunakan untuk meminimal  $J = \frac{1}{2} \|w\|^2$  dengan syarat (3) didapat persamaan (4)



Gambar 1. Ilustrasi pengklasifikasian data A dan B

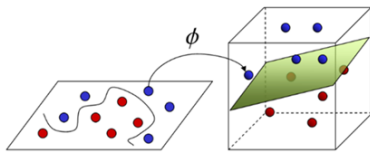
$$\bar{w} \cdot \bar{x}_i = \begin{cases} c_0 + a & \forall \bar{x}_i \in A \\ c_0 - a & \forall \bar{x}_i \in B \end{cases} \quad (1)$$

$$L = (\bar{x}_a - \bar{x}_b) \cdot \frac{\bar{w}}{\|w\|} = \frac{2a}{\|w\|} \quad (2)$$

$$b_i(\bar{w} \cdot \bar{x}_i - c_0) - a = 0 \quad (3)$$

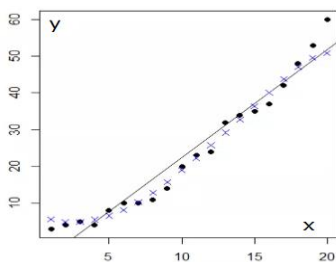
$$L = \sum_{i=1}^n \alpha_i a - \frac{1}{2} \sum_{i=1} \sum_{j=1} \alpha_i \alpha_j \beta_i \beta_j \bar{x}_i \cdot \bar{x}_j \quad (4)$$

Sehingga model Matematika SVM dapat ditentukan dari hasil perkalian titik pada setiap vektor datanya. Hal ini memungkinkan melakukan transformasi vektor  $\bar{x}_i$  ke bidang lain  $\phi(\bar{x}_i)$ . Transformasi vektor data ini dikenal dengan kernelisasi. Ilustrasi kernelisasi diperlihatkan pada Gambar 2.



Gambar 2. Proses kernelisasi dari 2D ke 3D dengan SVM

SVM dapat juga melakukan prediksi data seperti diperlihatkan pada Gambar 3. Sumbu x menyatakan variabel bebasnya sedangkan sumbu y menyatakan variabel terikatnya. Garis lurus merupakan hasil prediksi data menggunakan regresi linier sedangkan hasil regresi menggunakan SVM dinyatakan dalam tanda x yang jelas lebih baik dari prediksi regresi linier pada gambar.



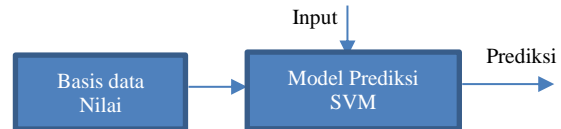
Gambar 3. Hasil Regresi linier dengan regresi SVM

SVM banyak diimplementasikan pada berbagai bidang. Pada bidang kedokteran untuk mendiagnosis tuberkulosis dari foto rontgen [3]. Fernandez [4] menggunakan SVM regresi untuk melakukan prediksi dinamis data deret waktu. Pratama dkk [5] memprediksi ketepatan waktu kelulusan mahasiswa. Yugopuspito dkk [6] menggunakan SVM untuk memprediksi posisi ikan tuna pada perairan Indonesia. Lukas dkk [7] melakukan

perhitungan asuransi keterlabatan pesawat terbang menggunakan SVM.

### 3. Perancangan Sistem

Sistem dirancang sebagai alat bagi mahasiswa untuk dapat menentukan bidang peminatan yang cocok berdasarkan minat mahasiswa dan kompetensi mahasiswa tersebut. Kompetensi mahasiswa diukur dari prestasi belajar di tahun pertama dan kedua. Perancangan sistem terdiri dari dua bagian, yaitu perancangan database dan pembentukan model prediksi SVM seperti diperlihatkan pada Gambar 4.



Gambar 4. Blok Diagram Sistem

Basis data nilai dipakai untuk membentuk model prediksi SVM. Basis data terdiri dari 80 responden dengan tiga peminatan yaitu *Interactive Media* (IM), *Medical Informatics* (MI) dan *Software Engineering* (SE). Setiap responden terdiri dari 21 matakuliah inti program studi hingga pada tahun kedua dengan total 67 sks. Lima matakuliah inti yang diajar di atas tahun ke dua dan sejumlah matakuliah wajib peminatan. Jumlah responden mahasiswa peminatan IM, MI dan SE masing-masing ada 30, 27 dan 23 mahasiswa. Indeks Prestasi Kumulatif setiap responden juga ada pada basis data nilai.

Fungsi prediksi untuk setiap bidang peminatan terdiri dari 21 variabel bebas yang merupakan nilai grade setiap matakuliah inti program studi di dua tahun pertama perkuliahan. Sedangkan variabel terikatnya adalah Index Prestasi kumulatif matakuliah peminatan. Matakuliah inti di dua tahun pertama diperlihatkan pada Tabel 1.

Tabel 1. Matakuliah inti Program Studi

| No | ID      | Nama Course                         | sks |
|----|---------|-------------------------------------|-----|
| 1  | INF0011 | Algoritma & Pemrograman             | 4   |
| 2  | INF0131 | Kalkulus 1                          | 3   |
| 3  | INF0221 | Matematika Diskrit                  | 4   |
| 4  | INF0381 | Pengantar Teknologi Informatika     | 3   |
| 5  | INF0141 | Kalkulus 2                          | 3   |
| 6  | INF0301 | Pemrograman Java                    | 4   |
| 7  | INF0531 | Sistem Informasi Manajemen          | 2   |
| 8  | INF0601 | Statistika                          | 3   |
| 9  | INF0621 | Struktur Data                       | 3   |
| 10 | INF0021 | Aljabar Linier & Matriks            | 2   |
| 11 | INF0031 | Analisis Algoritma                  | 2   |
| 12 | INF0191 | Komunikasi Data & Jaringan Komputer | 4   |
| 13 | INF0271 | Organisasi & Arsitektur Komputer    | 4   |
| 14 | INF0291 | Pemrograman Berorientasi Objek      | 3   |
| 15 | INF0521 | Sistem Basisdata                    | 4   |
| 16 | INF0081 | Grafika Komputer                    | 2   |
| 17 | INF0161 | Keamanan Komputasi                  | 4   |
| 18 | INF0281 | Pemodelan Berorientasi Objek        | 3   |
| 19 | INF0391 | Pengembangan Piranti Lunak          | 4   |
| 20 | INF0501 | Riset Operasional                   | 3   |
| 21 | INF0571 | Sistem Operasi                      | 3   |

#### 4. Hasil dan Pembahasan

Data penelitian terdiri dari tabel transaksi perkuliahan dan table matakuliah inti. Tabel perkuliahan mencatat kode mahasiswa, kode matakuliah dan grade matakuliah, sedangkan Tabel matakuliah inti mencatat kode matakuliah, nama matakuliah, semester dan sksnya. Dari dua file itu dibentuk satu data penelitian yang terdiri dari 21 field matakuliah inti di dua tahun pertama, 5 matakuliah inti di tahun selanjutnya dan 24 matakuliah pilihan untuk tiga peminatan, dan satu field IPK alumni. Kemudian ditambahkan dua field lainnya yaitu Peminatan dan IPK Inti.

Data pelatihan dibentuk dari data penelitian yang terdiri dari 80 record dengan 23 field. 21 field adalah nilai grade ke 21 matakuliah inti program studi dalam bentuk angka. Ke 21 field ini menjadi variable bebas sedangkan 2 field lainnya yaitu Peminatan dan IPK.Peminatan sebagai variable terikat. Distribusi grade dan nilai grade matakuliah diperlihatkan pada Tabel 2.

Tabel 2. Distribusi Nilai dan grade

| Score Nilai   | Grade | Nilai Grade |
|---------------|-------|-------------|
| 90.00 – 100   | A     | 4.0         |
| 85.00 – 89.99 | A-    | 3.7         |
| 80.00 – 84.99 | B+    | 3.3         |
| 75.00 – 84.99 | B     | 3.0         |
| 70.00 – 74.99 | B-    | 2.7         |
| 65.00 – 69.99 | C+    | 2.3         |
| 60.00 – 64.99 | C     | 2.0         |
| 55.00 – 59.99 | C-    | 1.7         |
| 40.00 – 54.99 | D     | 1           |
| 0.00 – 39.99  | E     | 0           |

Program disusun dalam Bahasa R yang diperlihatkan pada Algo 1 dan hasilnya pada Tabel 3. Terlihat bahwa keakurasian pemilihan peminatan mencapai 96.7%.

```
library (e1071)
Data_01 ← read.csv("Data.csv", header = TRUE)

var_X1 ← subset(Data_01, select=-Peminatan)
var_X1 ← subset(var_X1, select=-IPK.Peminatan)
var_X2 ← subset(Data_01,select=-IPK.Peminatan)
var_Y1 ← Data_01$Peminatan
var_Y2 ← Data_01$IPK.Peminatan

# Model
svm_minat ← svm(var_Y1 ~ ., data=var_X1)
summary(svm_minat)
prediction_minat ← predict(svm_minat, var_X1)
table(prediction_peminatan, var_Y1)

svm_IPK ← svm(var_Y2 ~ ., data=var_X2)
summary(svm_IPK)
pred_IPK ← predict(svm_IPK, var_X2)

Err2 = (var_Y2 - pred_IPK)^2
AccP = 1-abs(var_Y2 - pred_IPK)/var_Y2
Resu = cbind(Data_01, pred_IPK ,Err2,AccP)
write.csv(x = result, file = "hasil.CSV")
```

Algo 1. Pengklasifikasian jalur Peminatan

Tabel 3. Confusion Matrik Peminatan

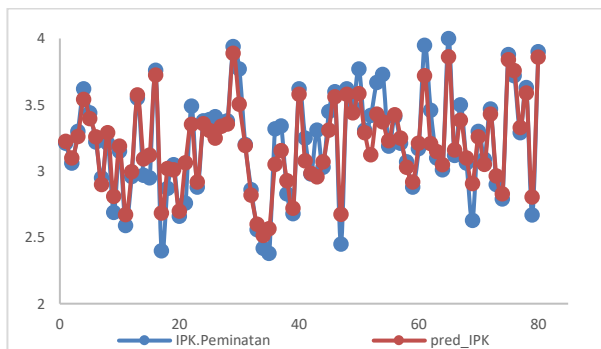
| Klasifikasi | IM | MI | SE |
|-------------|----|----|----|
| IM          | 30 | 0  | 0  |
| MI          | 0  | 26 | 1  |
| SE          | 0  | 1  | 22 |

Keakurasian pemilihan peminatan perlu dilengkapi dengan keakurasian prediksi IPK.Peminatan. Hasil prediksi IPK peminatan disimpan pada variable pred\_IPK, error kuadrat pada Err2 dan Akurasi prediksi IPK peminatan pada AccP. Tabel 4 memperlihatkan data statistik keempat variable itu.

Tabel 4.Rata-rata dan Standard deviasi IPK dan akurasinya

| Keterangan      | IPK  | Pred_IPK | Err2   | AccP   |
|-----------------|------|----------|--------|--------|
| Rata-rata       | 3.21 | 3.1977   | 0.0175 | 0.9692 |
| Standar deviasi | 0.40 | 0.3216   | 0.0293 | 0.0295 |

Pengujian hipotesis beda mean antara IPK.Peminatan dengan Pred.IPK memberikan hasil tidak cukup data untuk mengatakan rata-rata kedua data berbeda. Itu berarti dapat disimpulkan mereka mempunyai rata-rata yang sama. Pengujian perbedaan variance antara kedua data juga disimpulkan tidak cukup data untuk mengatakan variance kedua data berbeda. Ini berarti data IPK.Peminatan dengan prediksi IPK memiliki distribusi yang sama. Grafik IPK.Peminatan dengan Pred.IPK diperlihatkan pada Gambar 5.

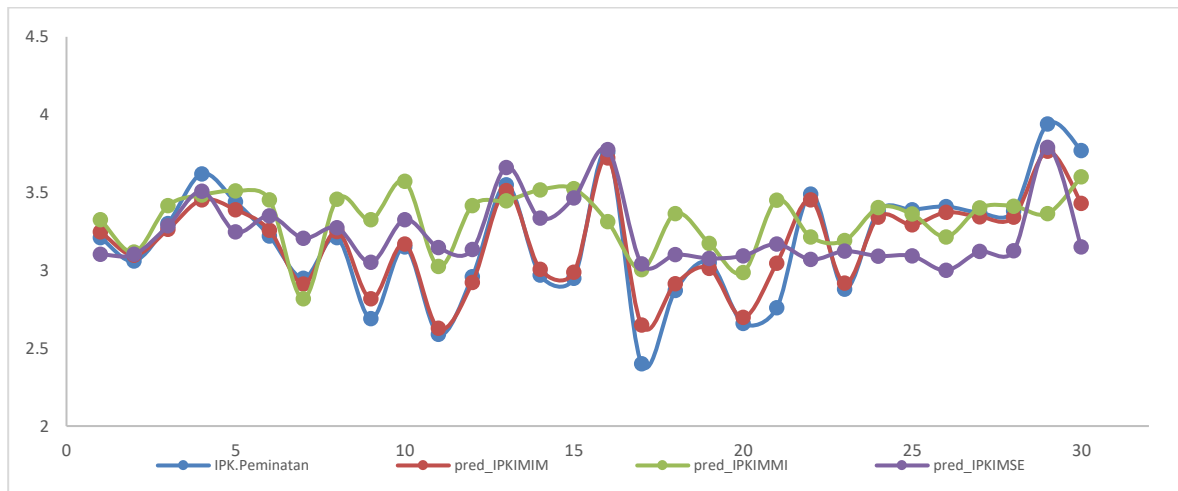


Gambar 5. Grafik IPK.Inti dengan Pred.IPK

Gambar 5 memperlihatkan hasil Prediksi IPK mengikuti IPK.Peminatan dengan simpangan yang relative kecil. Akurasinya mencapai 97% dengan standard deviasi 0.03. Hal ini mengindikasikan bahwa pred IPK memberikan hasil prediksi yang cukup baik terhadap IPK peminatannya.

Percobaan lanjutan adalah menguji berapa perbedaan rata-rata prediksi IPK peminatan A jika ia mengambil peminatan lainnya. Oleh karena itu, tiga model prediksi IPK peminatan dibentuk. Berdasarkan ketiga model prediksi peminatan ini dihitung prediksi IPK peminatan masing masing dan dihitung perbedaan rata-rata IPK peminatannya dengan IPK peminatan dasarnya. Hasil perhitungan untuk IPK dasar peminatan *Interactive media* diperlihatkan pada Gambar 6.

Hasil prediksi IPKIMIM artinya IPK peminatan m *Interactive Media* (IM) diprediksi IPK peminatannya dengan model prediksi IM. Grafik prediksi IPKIMIM mengikuti IPK Peminatannya namun tidak dengan prediksi IPKIMMI dan IPKIMSE. Rata-rata perbedaan IPKIMIM, IPKIMMI dan IPKIMSE terhadap IPKIM adalah 0.01, -0.15 dan -0.05. Ini artinya IPK peminatan IM akan meningkat sekitar 0.15 point jika ia mengambil peminatan MI dan 0.05 point untuk SE.



Gambar 6. Perbandingan Prediksi IPK Peminatan dengan IPK Peminatan Interactive Media

Rata-rata perbedaan IPKMIM, IPKMIMI dan IPKMISE terhadap IPKMI adalah 0.09, -0.02 dan 0.02. Sedangkan rata-rata perbedaan IPKSEIM, IPKSEMI dan IPKSESE terhadap IPKSE adalah 0.00, -0.07 dan 0.00. Ini artinya IPK peminatan IM ataupun SE tidak berubah jika ia ke peminatan lainnya.

#### 4. Kesimpulan

Tiga model prediksi IPK peminatan yaitu *Interactive Media*, *Medical Informatics* dan *Software Engineering* telah berhasil dikembangkan dengan tingkat ketepatan mencapai 97 % dari data pelatihan. Namun prediksi IPK peminatan atas input variable bebas prestasi belajar di dua tahun pertama sebanyak 21 matakuliah inti program studi memberikan hasil ketepatan mencapai 99 persen. Ada perbedaan prediksi sebesar 0.15 point dari peminatan *Interactive Media* ke *Medical informatic*.

#### Daftar Rujukan

[1] Samuel Lukas, Meiliyana, William Simson, 2009. Penerapan Logika Fuzzy dalam pengambilan Keputusan untuk Jalur Peminatan Mahasiswa. Kampus Renon, STMIK STIKOM Bali,

*Konferensi Nasional Sistem & Informatika 2009*, 14 November, 2009, Bali, Indonesia.

- [2] Kecman V., 2014. *Support Vector Machines – An Introduction*. The University of Auckland, School of Engineering, Auckland, New Zealand.
- [3] Veropoulos K., Cristianini N., and Campbell C., 1999. *The Application of Support Vector Machines to Medical Decision Support: A Case Study*, Department of Engineering Mathematics, Bristol University, Bristol BS8 1TR, United Kingdom.
- [4] Fernandez R., 1999. *Predicting Time Series with a Local Support Vector Regression Machine*. Institut Galilee, Paris, Prancis.
- [5] Arif Pratama, Randy Cahya Wihandika, Dian Eka Ratnawati, 2018. Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 4. April 2018, pp. 1704-1708
- [6] Yugopuspito P., Samuel Lukas, I Made Murwantara, Steven Albert, Hery, 2019. A Novel Fishing Finder Technique Based on VMS Data in Indonesia. *ACM Conference Proceedings The 3rd International Conference on Advances in Artificial Intelligence (ICAAI 2019)*, Bahcesehir University, Istanbul, Turkey. pp. 107-111
- [7] Lukas S., Dina Stefani, Petrus W., 2019. Comparing SVM and GLM in Calculating Insurance Premium for Flight Delay. *ACM Conference Proceedings The 3rd International Conference on Advances in Artificial Intelligence (ICAAI 2019)*, Bahcesehir University, Istanbul, Turkey. pp. 141-145