

## **Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus SMKN 1 Sutera)**

Rycki Rahmatil Fiska

*Program Studi Teknik Komputer, AMIK Mitra Gama, Riau, Indonesia*

*email: rahmatilrf.09@gmail.com*

### **Abstrak**

*Drop out adalah suatu keadaan di mana siswa diberhentikan dari sekolah karena beberapa alasan yang mengharuskan siswa untuk di drop out . faktor drop out seperti absen yang telah melebihi batas maksimal, nilai yang kurang dari batas bawah yang telah ditentukan pihak sekolah dan perilaku siswa yang sudah melanggar nilai attitude yang telah diterapkan oleh pihak sekolah. Dalam penelitian ini penulis ingin memprediksi siswa yang berpotensi drop out dengan menggunakan metode SVM dan diharapkan mampu memprediksi masalah drop out apakah siswa akan lanjut ke semester selanjutnya atau di drop out karena beberapa alasan. Dan hasil yang disimpulkan adalah metode SVM mampu memprediksi siswa yang berpotensi drop out.*

**Kata kunci :** *Support Vector Machine, Kernel Trick, Drop Out.*

### **1. Pendahuluan**

*Drop Out* adalah keluar dari sekolah sebelum waktunya, atau sebelum lulus. *Drop out* demikian ini perlu dicegah, oleh karena hal demikian dipandang sebagai pemborosan bagi biaya yang sudah terlanjur dikeluarkan untuknya. Banyaknya peserta didik yang *drop out* adalah indikasi rendahnya produktivitas pendidikan. Tingginya angka *drop out* juga bisa mengganggu angka partisipasi pendidikan atau sekolah.

Pada umumnya di sekolah-sekolah sekarang ini dibedakan 3 hal sehubungan dengan masalah ketidakhadiran. Penyebab ketidakhadiran tersebut diantaranya adalah adanya ijin, sakit dan alpa. Tetapi ketiga hal tersebut akan menyebabkan sebuah masalah jika dalam jumlah yang sering dilakukan oleh peserta didik. Salah satu akibat yang akan diterima oleh peserta didik adalah sebuah pilihan yang harus diterima yaitu sebuah pernyataan *drop out* dari sekolah. Secara umum sebab-sebab terjadinya *drop out* yaitu peserta didik tidak mampu menyelesaikan pendidikan, tidak mempunyai biaya sekolah, peserta didik dalam keadaan sakit dan tidak kunjung sembuh.

SVM merupakan metode yang berkaitan dengan *Data Mining* karena SVM merupakan algoritma klasifikasi berdasarkan

prinsip *linear classifier*. SVM telah banyak digunakan diberbagai penelitian. Salah satu masalahnya adalah Aplikasi Metode Cross Entropy Untuk Support Vector Machines yang merupakan penelitian dari Tiananda Widyarini dan Budi Santosa Jurusan Teknik Industri Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pada penelitian ini, metode CE diterapkan untuk menyelesaikan masalah optimasi *lagrange* SVM agar didapatkan proses komputasi yang lebih cepat dan sederhana. Metode yang telah dikembangkan kemudian diuji pada enam dataset kasus nyata. Dari hasil pengujian, didapatkan kesimpulan bahwa aplikasi CE pada SVM mampu mengklasifikasi data dua kelas dengan akurasi yang sebanding dengan metode SVM standar. Sebagai tambahan, metode ini mampu menyelesaikan permasalahan dengan waktu komputasi yang lebih cepat daripada SVM standar untuk data berukuran besar.

### **2. Landasan Teori**

#### **2.1. Data Mining**

*Data Mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. *Data Mining* merupakan proses semi otomatis yang

menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar (Turban et al, 2005).

*Data Mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (Angga & Riani, 2012).

Menurut Gartner Group *Data Mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005). Selain definisi diatas definisi juga diberikan seperti tertera dibawah ini :

*Data Mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006).

*Data Mining* merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, *forecasting*, *asociation rule*, *sequential pattern*, *clustering*, *regression*, deskripsi dan visualisasi.

## 2.2. Support vector machines (SVM)

*Support vector machines* (SVM) adalah suatu metode yang handal dalam menyelesaikan masalah klasifikasi data. Permasalahan SVM dipecahkan dengan menyelesaikan persamaan *Lagrangian* yang merupakan bentuk dual dari SVM melalui *quadratic programming* (Tiananda, Budi, 2009).

*Support Vector Machine* (SVM) diusulkan sebagai alternatif dari SVM standar yang telah terbukti lebih efisien daripada SVM tradisional dalam pengolahan data berskala besar (Chia Hui Huang-2014).

Klasifikasi adalah tugas *Data Mining* yang memprediksi pengelompokan untuk kasus data. Dalam aplikasi bidang pendidikan dari metode klasifikasi. Contohnya mengelompokan karya siswa, predikat nya dan nilai akhir nya (Sajadin, dkk, 2011).

Selanjutnya SSVM menghasilkan dan memecahkan reformulasi mulus tanpa hambatan untuk klasifikasi pola yang menggunakannya. Algoritma SSVM dipecahkan sangat cepat oleh Newton - Armijo dan telah diperpanjang untuk *furfaces* pemisahan *non linear* dengan menggunakan teknik kernel *non linear*. Hasil penelitian ini menunjukkan *numeric* SSVM lebih cepat dari metode lain dan memiliki kemampuan generalisasi yang lebih baik (Sajadin, dkk, 2011).

Penelitian dilakukan melalui simulasi komputer menggunakan model SVM linear dan SVM non linear. Berikut ini akan diuraikan secara ringkas tentang ide dasar SVM. Misalkan diberikan himpunan  $X = \{x_1, x_2, \dots, x_m\}$ , dengan  $X_i \in R^n$ ,  $i=1, \dots, m$ , telah diketahui bahwa  $X$  berpolat tertentu, yaitu apabila  $x_k$  termasuk dalam suatu kelas maka  $x_k$  diberikan label (merupakan target)  $y_k = -1$ . Dengan demikian data yang diberikan berupa pasangan  $(x_1, y_1), \dots, (x_m, y_m) \in X \times \{-1, 1\}$ . Dalam masalah pembelajaran, kumpulan pasangan tersebut merupakan data pembelajaran SVM. Berbekal pengalaman pembelajaran menggunakan data pembelajaran tersebut, SVM harus mampu menentukan pola (generalisasi) dari  $x \in X$ .

$$X_n = \frac{0.8 * (X - a)}{b - a} + 0.1$$

(1)

Di mana :

$X_n$	=	Nilai ke - n
A	=	Nilai angka terendah
B	=	Nilai angka tertinggi
0.8 dan 0.1	=	Ketetapan

Dengan menggunakan persamaan tersebut, maka kita dapat mencari nilai dari transformasi data  $x_1$  (Kehadiran) dan  $x_2$  (Nilai).

Masalah dasar dari SVM adalah menentukan suatu *hyperplane*  $\langle w, x \rangle + b = 0$  memisahkan data  $x_j$  yang terdiri dari dua kelas, yaitu  $y_i = \{+1, -1\}$ , dengan *margin* maksimal. *Margin* disini merupakan jarak antara *hyperplane* ke masing-masing kelas data. Selanjutnya, *hyperplane* ini akan menjadi fungsi keputusan  $f(x)$  untuk masalah klasifikasi dua kelas diatas.

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x) + b) =$$

$$\text{sign} \left( \sum_{i=1}^N \alpha_i y_i \phi(x_i)^T \cdot \phi(x) + b \right) \quad (2)$$

Dimana :

$w$  = nilai *weight*  
 $x$  = nilai variabel input  
 $b$  = nilai *bias*

Rumus di atas digunakan untuk perhitungan hasil prediksi.

*Hyperplane* tersebut dapat dibuat secara tunggal sesuai dengan  $w$  dan  $b$  yang akan diperoleh. Selanjutnya  $x_i$  yang berupa subhimpunan data pembelajaran yang terletak pada *margin* disebut dengan *support vector*. Untuk mencari nilai matrik  $K$  menggunakan persamaan seperti berikut :

$$K(x, x_i) = \phi(x_i) \quad (3)$$

Menurut teorema Vapnik, pemilihan *margin* maksimum akan diberikan kemampuan generalisasi yang baik. Untuk menentukan persamaan *hyperplane*  $\langle w, x \rangle + b = 0$  harus diketahui lebih dahulu nilai  $w$  dan  $b$ . Disini nilai  $w$  merupakan nilai yang berkaitan dengan *margin*. Untuk memperolehnya digunakan bantuan *hyperplane* kanonik yaitu :

Dalam hal ini,  $x^+$  adalah data yang terletak pada kelas  $y = +1$  dan terdekat ke *hyperplane* dan  $x^-$  adalah data yang terletak pada kelas  $y = -1$  dan terdekat ke *hyperplane*. Lebar *margin*  $y$  adalah jarak  $x^+$  ke *hyperplane* atau jarak  $x^-$  ke *hyperplane*. Jadi memaksimalkan *margin* adalah ekuivalen dengan memaksimalkan  $\|w\|$  dengan syarat:

*Hyperplane*  $w$  dan  $b$  merupakan penyelesaian dari PK, sebagai berikut :

$$w = \sum_{i=1}^N \alpha_i y_i K(x, x_i)$$

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-)$$

Dengan kendala :  $y_i(\langle w, x_i \rangle + b) \geq 1, i = 1 \dots m$  adalah merupakan *hyperplane* dengan *margin* maksimum, yaitu  $y = 1/\|w\|$ . Bentuk primal PK persamaan dapat di bawa kedalam fungsi *langrange* sebagai berikut dengan  $\alpha_i \geq 0$  adalah pengali *langrange*.

Nilai bias dapat dicari dari nilai kendala PK pada posisi 1. Sedangkan  $b^*$  merupakan  $b$  optimal menentukan penggeseran *hyperplane* ke kiri ( data maksimal pada kelas  $-1$ ) dan kekanan ( data minimal ke kelas  $+1$ ) yang paling jauh. Oleh karena itu tujuan dan kendala dari PK adalah berupa fungsi konveks, maka penyelesaian PK tersebut merupakan

penyelesaian optimal *global*. Persamaan untuk mencari nilai alpha adalah :

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \frac{6}{\sum K(N \times N)} \quad (6)$$

Dalam menggunakan model SVM, data yang digunakan dipilih menjadi data pembelajaran, data validasi dan data uji. Data pembelajaran digunakan dalam membentuk SVM, sementara nilai parameter bebasnya dipilih dari nilai parameter yang membuat galat dari data validasi bernilai minimal. Selanjutnya data SVM yang dihasilkan digunakan untuk memprediksi data uji. Untuk fungsi *kernel* yang dapat dipakai pada SVM adalah sebagai berikut :

1. *Polynomial* ( $X^T X_i + 1$ )<sup>p</sup>
2. *Radial Basic Funtion* ( *RBF* ) :  $\exp ( - 1/2a^2 \| x - x_i \|^2 )$

### 3. Metodologi Penelitian

Agar langkah-langkah yang diambil peneliti dalam penelitian dapat lebih mudah dipahami, berikut ini disertakan penjelasan terperinci mengenai sistematika model metodologi pemecahan masalah.

#### 1. Pengumpulan Data / informasi

Pengumpulan data atau informasi merupakan tahap awal dalam metodologi penelitian. Pada tahap ini dilakukan pengumpulan data dan informasi dari yang diperoleh langsung dengan melakukan wawancara kepada petugas SMK N 1 Sutera.

Kemudian pengumpulan data<sup>(4)</sup> serta informasi juga diperoleh dari buku-buku dan jurnal-jurnal serta dari berbagai literatur yang mendukung dalam proses penelitian ini<sup>(5)</sup>. Hal tersebut dijadikan sebagai landasan pemikiran dalam penelitian.

#### 2. Penganalisaan Data dan Masalah

Proses penganalisaan masalah dilakukan untuk mendapatkan fakta yang dapat digunakan untuk memecahkan masalah yang sedang diteliti. Analisa digunakan untuk melihat masalah yang sedang terjadi yang akan digunakan sebagai bahan yang akan diteliti. Metode yang digunakan dalam penelitian ini adalah *Data Mining* dengan menggunakan metode SVM (*Support Vector Machine*). Di mana metode SVM (*Support Vector Machine*) digunakan sebagai teknik pembelajaran pada *Data Mining* tersebut.

3. Identifikasi Dan Perumusan Masalah

Setelah dilakukan kedua tahap sebelumnya, barulah kemudian dilakukan identifikasi masalah serta perumusan masalah. Masalah yang didapatkan adalah bagaimana menganalisa prediksi siswa yang berpeluang *drop out* dengan menggunakan *Data Mining SVM (Support Vector Machine)*. Oleh karena itu penulis ingin menganalisa prediksi siswa yang berpeluang *drop out* dengan menggunakan *Data Mining SVM (Support Vector Machine)* pada SMK N 1 Sutera

4. Studi Kepustakaan

Setelah masalah diperoleh serta telah dianalisa maka akan dilakukan studi kepustakaan yang berguna untuk memperoleh referensi serta literatur dalam menyelesaikan masalah. Studi kepustakaan ini dilakukan dengan cara mencari referensi berupa buku atau jurnal, baik dengan melalui media *internet* yang berkaitan dengan *SVM (Support Vector Machine)* serta tentang prediksi siswa yang berpeluang *drop out*.

5. Analisa SVM (*Support Vector Machine*)

Setelah referensi diperoleh maka selanjutnya dilakukan analisa terhadap *SVM* yaitu mempelajari rumus-rumus dan mulai menghubungkan data yang diperoleh dengan rumus-rumus tersebut sehingga data tersebut menjadi data baru yang akan menghasilkan perhitungan terhadap siswa yang berpeluang *drop out*.

6. Implementasi (Pelatihan)

Pada tahapan implementasi bertujuan untuk mendapatkan hasil dari analisa tentang prediksi siswa yang *drop out*. Setelah data dianalisa kemudian dilakukan pelatihan untuk mendapatkan hasil yang diinginkan. Pelatihan dilakukan dengan menggunakan dua tahapan. Tahapan – tahapan yang digunakan dalam pelatihan yaitu :

1. Pelatihan secara manual

Pelatihan secara manual dilakukan dengan mengolah data yang telah didapatkan dari proses analisa manual. Kemudian data tersebut di masukkan ke dalam rumus – rumus perhitungan metode *SVM (Support Vector Machine)*.

2. Pelatihan secara *software*

Data yang telah di analisa akan dilakukan pelatihan secara *software*. Pelatihan secara *software* dilakukan dengan menggunakan *software Matlab 6.1*. Hal ini sangat berpengaruh terhadap penelitian, oleh karena itu data tersebut harus akurat agar dapat digunakan untuk

menganalisa peluang siswa yang *drop out* di SMK N 1 Sutera. Tahapan implementasi akan dijelaskan pada BAB IV dan BAB V dalam laporan penelitian ini.

7. Pengujian

Pada tahapan pengujian dilakukan untuk mengetahui cara sistem bekerja dalam mengolah data. Proses yang dilakukan yaitu mengolah data mentah untuk menjadi data baru dengan menggunakan rumus-rumus *SVM* dan dilakukan perhitungan secara manual. Data mentah tersebut yaitu data absensi dan data nilai siswa berupa nilai akhir dari semester awal sampai semester yang sedang berlangsung pada siswa kelas XI jurusan listrik di SMK N 1 Sutera. Data tersebut akan diuji dengan *software Matlab 6.1*. Proses tersebut dilakukan untuk mengetahui apakah ada perbedaan atau tidak.

## 4. Analisa Dan Perancangan

Dalam menganalisa suatu masalah baik prediksi maupun klasifikasi, terlebih dahulu kita harus menentukan variabel dari data tersebut. Variabel dari data penyebab *drop out* siswa terdiri dari nilai, faktor kehadiran, perilaku, mata pelajaran, guru, biaya sekolah dan banyak lainnya. Dari variabel tersebut kita akan memilih variabel yang berpengaruh terhadap masalah *drop out* yaitu nilai dan kehadiran. Nilai diperoleh dari data nilai siswa yang ada pada SMK N 1 Sutera. Dan kehadiran berdasarkan dari jumlah wajib hadir siswa dalam satu semester yaitu 96 hari setelah dikurangi hari libur nasional dan waktu ujian. Jika absen siswa sebanyak 9 hari dalam satu semester maka kehadiran menjadi jumlah wajib hadir siswa dikurangi banyak absen yaitu :  $96 - 9 = 87$  dan begitu juga untuk selanjutnya.

Kita dapat mencari nilai dari transformasi data  $x_1$ (Kehadiran) dan  $x_2$ (Nilai) dengan menggunakan persamaan no (1) dengan cara cara:

$$X_n = \frac{0,8 * (87 - 81)}{96 - 81} + 0,1 = \frac{4,8}{15} + 0,1 \\ = 0,32 + 0,1 = 0,42$$

$$X_n = \frac{0,8 * (1591 - 1208)}{1775 - 1208} + 0,1 \\ = \frac{306,4}{567} + 0,1$$

$$= 0,54 + 0,1 = 0,64$$

Dimana :

0.8 = ketetapan rumus

0.1 = ketetapan rumus

87 = jumlah kehadiran pada data pertama

81 = jumlah kehadiran terendah

96 = jumlah kehadiran tertinggi

1591 = jumlah nilai pada data pertama

1208 = jumlah nilai terendah

1775 = jumlah nilai tertinggi

Sehingga diperoleh transformasi data  
 $x1(\text{Kehadiran}) = 0.42$ , dan transformasi data  
 $x2(\text{Nilai}) = 0.64$ . Lakukan langkah yang sama  
 untuk seluruh data.

**Tabel 1. Transformasi Data**

X1	X2	X3
0.42	0.74	0.64
0.69	0.26	0.74
0.85	0.42	0.9
0.1	0.42	0.1
0.63	0.58	0.61
0.26	0.1	0.31
0.37	0.42	0.62
0.15	0.42	0.58
0.85	0.58	0.67
0.74	0.74	0.61
0.58	0.74	0.7
0.85	0.74	0.65
0.9	0.42	0.83
0.42	0.42	0.6
0.47	0.42	0.59
0.58	0.74	0.65
0.74	0.1	0.78
0.31	0.26	0.41
0.74	0.9	0.47
0.74	0.42	0.63
0.85	0.1	0.64
0.31	0.26	0.59
0.85	0.42	0.7
0.9	0.42	0.9
0.63	0.58	0.36
0.15	0.58	0.62
0.37	0.58	0.4
0.69	0.42	0.27
0.74	0.58	0.65
0.85	0.58	0.61
0.85	0.74	0.83
0.26	0.26	0.17
0.69	0.26	0.57
0.63	0.42	0.82

0.85	0.42	0.79
0.79	0.74	0.65

Berdasarkan data pada tabel 4.2 terlihat bahwa untuk nilai X1 (1) 87 adalah 0.42, X2 (1) 90 adalah 0.74 dan X3 (1) 1591 adalah 0.64.

Untuk X1, X2, dan X3 diambil dari data yang telah ditransformasikan yaitu urutan yang ke 21 sampai data urutan 26. Untuk X1 adalah = 0.84, 0.31, 0.84, 0.9, 0.63, 0.15. untuk X2 adalah = 0.63, 0.58, 0.70, 0.9, 0.35, 0.62. dan untuk hasil prediksi sementara X2-X1 (0.63 - 0.84 = -0.21). jika hasil prediksi sementara adalah negatif maka label data latih (-1), jika hasil prediksi sementara positif maka label data latih (+1).

**Tabel 2. Data Latih**

X1	X2	X3	Label
0.42	0.74	0.64	-1
0.69	0.26	0.74	1
0.1	0.42	0.1	-1
0.26	0.1	0.31	1
0.85	0.58	0.67	-1
0.42	0.42	0.6	1

**Tabel 3. Data Uji**

X1	X2	X3	Label
0.85	0.42	0.9	1
0.63	0.58	0.61	-1
0.37	0.42	0.62	1
0.15	0.42	0.58	1
0.74	0.74	0.61	-1
0.58	0.74	0.7	-1
0.85	0.74	0.65	-1
0.9	0.42	0.83	-1
0.47	0.42	0.59	1
0.58	0.74	0.65	-1
0.74	0.1	0.78	1
0.31	0.26	0.41	1
0.74	0.9	0.47	-1
0.74	0.42	0.63	-1
0.85	0.1	0.64	-1
0.31	0.26	0.59	1
0.85	0.42	0.7	-1
0.9	0.42	0.9	1
0.63	0.58	0.36	-1
0.15	0.58	0.62	1
0.37	0.58	0.4	-1
0.69	0.42	0.27	-1
0.74	0.58	0.65	-1
0.85	0.58	0.61	-1

0.85	0.74	0.83	-1
0.26	0.26	0.17	-1
0.69	0.26	0.57	-1
0.63	0.42	0.82	1
0.85	0.42	0.79	-1
0.79	0.74	0.65	-1

Pada penelitian ini penulis mempunyai 36 data. Setelah menghitung nilai matrik kernel  $K$ , maka didapatkan matrik  $K$  dengan ukuran  $N \times N$ .

$$\begin{pmatrix} 4.552 & 3.825 & 2.007 & 1.908 & 4.906 & 3.501 \\ 3.825 & 4.373 & 1.568 & 2.058 & 4.986 & 3.396 \\ 2.007 & 1.568 & 1.431 & 1.207 & 1.947 & 3.396 \\ 1.908 & 2.058 & 1.207 & 1.377 & 2.210 & 1.788 \\ 4.906 & 4.986 & 1.947 & 2.210 & 6.289 & 4.010 \\ 3.501 & 3.396 & 1.634 & 1.788 & 4.010 & 2.933 \end{pmatrix}$$

Karena pada fitur label memiliki 6 buah data, maka untuk  $\alpha$  juga memiliki 6 data yang sama sehingga setiap elemen matrik kernel  $K(N \times N)$  digunakan untuk menghitung nilai  $\alpha$  di mana jumlah  $\alpha$  dibagi dengan jumlah kernel  $K(N \times N)$  yang dikali dengan  $y_i y_j$ .

Maksudnya untuk mendapatkan nilai  $\alpha$  kita harus menjumlahkan semua nilai pada matrik kernel  $K$  dengan berpedoman pada label data latih pada tabel 4.3 yaitu jika label data latihnya (-1) maka nilai matriknya juga (-) dan jika label data latihnya (+) maka nilai matriknya juga (+). Persamaan ini untuk masing-masing kolom matrik. Maka nilai  $\alpha$  dengan persamaan no (6) adalah :

$$a1 = a2 = a3 = a4 = a5 = a6 = \frac{6}{3.4609} = 1.73365$$

Setelah didapatkan nilai  $\alpha$ , langkah selanjutnya adalah mencari nilai  $weight$  ( $w$ ) dan  $bias$  ( $b$ ) dengan persamaan no (4). Artinya nilai  $\alpha$  di kali dengan nilai matrik per kolom. Nilai  $\alpha$  (+) atau (-) masih berpedoman pada label data latih.

$$= \begin{bmatrix} -3.8665 \\ -0.9555 \\ -1.6925 \\ -0.1777 \\ -3.3556 \\ -1.7817 \end{bmatrix}$$

Setelah didapatkan nilai  $w$ , selanjutnya pilih salah satu *Support Vector* dari kelas “+1” dan “-1” dari matrik yang mempunyai nilai (+)

dan (-). Pada penelitian ini di ambil dari kolom pertama dan kolom ke enam untuk menghitung nilai  $b$  dengan persamaan no (5) :

$$= -\frac{1}{2} * \begin{bmatrix} 4.5522 \\ 3.8252 \\ 2.0073 \\ 1.9088 \\ 4.9062 \\ 3.5014 \end{bmatrix} \begin{bmatrix} -3.8665 \\ -0.9555 \\ -1.6925 \\ -0.1777 \\ -3.3556 \\ -1.7817 \end{bmatrix} + \begin{bmatrix} 3.5014 \\ 3.3966 \\ 1.6343 \\ 1.7881 \\ 4.0104 \\ 2.9337 \end{bmatrix} \begin{bmatrix} -3.8665 \\ -0.9555 \\ -1.6925 \\ -0.1777 \\ -3.3556 \\ -1.7817 \end{bmatrix} = 43.1231$$

Dari hasil matrik  $K$  di atas, selanjutnya akan dicari nilai dari  $f(\phi(x))$ , di mana hasil dari matrik  $K$  dikalikan dengan  $weight$ . Dari hasil perkalian tersebut akan didapatkan nilai  $w * \phi(x)$  yang selanjutnya ditambahkan dengan nilai  $bias$  ( $b$ ) dengan menggunakan persamaan no (2). Artinya untuk mendapatkan nilai  $f(\phi(x))$ , kita harus mendapatkan hasil dari model SVM di kali nilai  $weight$  dan di tambah dengan nilai  $bias$ . Seperti di bawah ini :

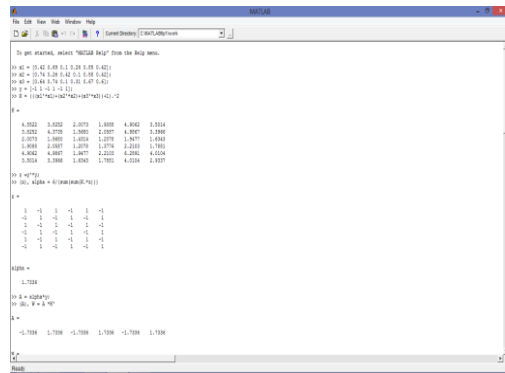
$$= \text{sign} (((-19.4663) + (-5.3294) + (-3.091) + (-0.4225) + (-22.148) + (-7.6595)) + 43.1231)) = \text{sign} (-14.9936) = -1$$

## 5. Implementasi dan Hasil

### 1. Mencari nilai kernel $K(N \times N)$

Untuk mencari nilai kernel  $K(N \times N)$ , terlebih dahulu kita harus memasukkan variable-variabel yang dibutuhkan dalam memproses data prediksi. Variabel yang diperlukan adalah :

```
>> x1 = [0.42, 0.69, 0.1, 0.26, 0.85, 0.42];
>> x2 = [0.74, 0.26, 0.42, 0.1, 0.58, 0.42];
>> x3 = [0.64, 0.74, 0.1, 0.31, 0.67, 0.6];
>> y = [-1 1 -1 1 -1 1];
```



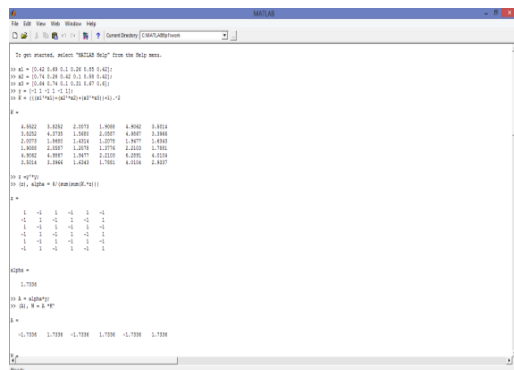
Gambar 1. Nilai Kernel K(N\*N)

Berikut adalah nilai Kernel K(N\*N) data latih dan nilai  $\alpha$ .

```
>> K = (((x1'*x1)+(x2'*x2)+(x3'*x3))+1).^2
K =
```

4.552	3.825	2.007	1.908	4.906	3.501
3.825	4.373	1.568	2.058	4.986	3.396
2.007	1.568	1.431	1.207	1.947	3.396
1.908	2.058	1.207	1.377	2.210	1.788
4.906	4.986	1.947	2.210	6.289	4.010
3.501	3.396	1.634	1.788	4.010	2.933

## 2. Pencarian Nilai $\alpha$



Gambar 2. Mencari Nilai  $\alpha$

Nilai  $\alpha$  yang diperoleh adalah 1.7336. berikut adalah langkah-langkah untuk mencari nilai  $\alpha$  menggunakan Matlab 6.1.

```
>> z = y'*y;
```

Nilai "z" merupakan hasil dari perkalian nilai  $y_i y_j$ .

```
>> (z), alpha = 6/(sum(sum(K.*z)))
```

1	-1	1	-1	1	-1
-1	1	-1	1	-1	1
1	-1	1	-1	1	-1
-1	1	-1	1	-1	1
1	-1	1	-1	1	-1

-1	1	-1	1	-1	1
----	---	----	---	----	---

$\alpha = 1.7336$

## 3. Pencarian Nilai $Weight$

Agar lebih memahami kalkulasi data pada software berikut adalah penjelasan mengenai langkah-langkah pencarian nilai  $weight$  pada Matlab 6.1.

```
>> A = alpha*y;
```

```
>> (A), W = A *K'
```

$A = -1.733 \quad 1.733 \quad -1.733 \quad 1.733 \quad -1.733 \quad 1.733$

$W = -3.866 \quad -0.955 \quad -1.692 \quad -0.177 \quad -3.355$   
 $-1.781$

## 4. Pencarian Nilai $bias$

Berikut adalah langkah-langkah yang dilakukan dalam pencarian nilai  $bias$  menggunakan Matlab 6.1.

```
>> (W),(K(1,:)),(K(6,:)), b = (-0.5)*(sum(W.*K(1,:))+sum(W.*K(6,:)))
```

$W =$

$-3.8667 \quad -0.9553 \quad -1.6925 \quad -0.1778$   
 $-3.3555 \quad -1.7816$

$ans = 4.5522 \quad 3.8252 \quad 2.0073 \quad 1.9088$   
 $4.9062 \quad 3.5014$

$ans = 3.5014 \quad 3.3966 \quad 1.6343 \quad 1.7881$   
 $4.0104 \quad 2.9337$

$b = 43.1227$

## 5. Melakukan Pengujian dengan Data Uji

Setelah nilai Kernel, nilai  $Weight$ , dan nilai  $bias$  diperoleh, Maka kita akan melakukan pengujian terhadap data uji yang berjumlah 30 data. Dalam pencarian nilai kernel  $f(\phi(x))$ , langkah awal yang harus dilakukan adalah memasukkan variabel data uji yang terdapat pada tabel 4.4 pada bab sebelumnya di mana  $X_1$ ,  $X_2$  dan  $X_3$  pada data uji dikalikan dengan  $X_1$ ,  $X_2$  dan  $X_3$  pada data latih. Variabel yang dibutuhkan adalah :

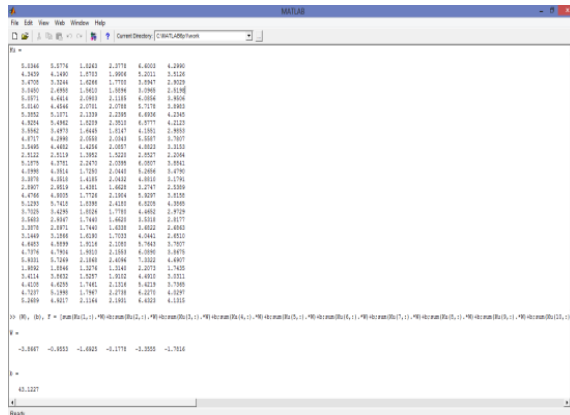
$xu = [0.85 \quad 0.63 \quad 0.37 \quad 0.15 \quad 0.74 \quad 0.58 \quad 0.85 \quad 0.9$   
 $0.47 \quad 0.37 \quad 0.15 \quad 0.58 \quad 0.85 \quad 0.58 \quad 0.74 \quad 0.31 \quad 0.74$   
 $0.74 \quad 0.85 \quad 0.31 \quad 0.85 \quad 0.9 \quad 0.63 \quad 0.15 \quad 0.37 \quad 0.69$   
 $0.63 \quad 0.85 \quad 0.79];$

$yu = [ \quad 0.42 \quad 0.58 \quad 0.42 \quad 0.42 \quad 0.74 \quad 0.74 \quad 0.74$   
 $0.42 \quad 0.42 \quad 0.74 \quad 0.1 \quad 0.26 \quad 0.9 \quad 0.42 \quad 0.1 \quad 0.26$



0.42 0.42 0.58 0.58 0.58 0.42 0.58 0.58 0.74  
0.26 0.26 0.42 0.42 0.74];

zu = [ 0.9 0.61 0.62 0.58 0.61 0.7 0.65 0.83  
0.59 0.65 0.78 0.41 0.47 0.63 0.64 0.59 0.7  
0.9 0.36 0.62 0.4 0.27 0.65 0.61 0.83 0.17  
0.57 0.82 0.79 0.65];



Gambar 3. Nilai kernel K dari Data Uji

Berikut adalah langkah-langkah yang diperlukan untuk mencari nilai kernel pada data uji.

```
>> Ku = ((xu'*x1)+(yu'*x2)+(yu'*x3)+1).^2
```

Ku =

5.0346	5.5776	1.8263
2.3778	6.6003	4.2990
4.3439	4.1490	1.8703
1.9906	5.2011	3.5126
3.4708	3.3244	1.6266
1.7700	3.8947	2.9029
3.0450	2.6958	1.5610
1.5896	3.0965	2.5193
5.0571	4.6414	2.0903
2.1185	6.0856	3.9506
5.0140	4.4546	2.0701
2.0788	5.7178	3.8983
5.3852	5.1071	2.1339
2.2395	6.6936	4.2345
4.9284	5.4962	1.8209
2.3510	6.5777	4.2123
3.5562	3.4973	1.6445
1.8147	4.1551	2.9853
4.8717	4.2998	2.0558
2.0343	5.5557	3.7807
3.5495	4.4682	1.4256
2.0857	4.8823	3.3153
2.5122	2.5119	1.3952
1.5220	2.8527	2.2064
5.1875	4.3751	2.2470
2.0395	6.0807	3.8841

4.0998	4.3514	1.7250
2.0440	5.2656	3.4790
3.3878	4.3518	1.4185
2.0432	4.8810	3.1791
2.8907	2.9519	1.4381
1.6628	3.2747	2.5389
4.4766	4.9005	1.7726
2.1904	5.9597	3.8158
5.1293	5.7418	1.8398
2.4180	6.8205	4.3865
3.7025	3.4295	1.8026
1.7780	4.4652	2.9729
3.5633	2.9347	1.7440
1.6620	3.5318	2.8177
3.3878	2.8971	1.7440
1.6620	3.5318	2.8177
4.6483	4.5899	1.9116
2.1080	5.7643	3.7807
4.7376	4.7904	1.9310
2.1553	6.0890	3.8675
5.9331	5.7269	2.1863
2.4096	7.3322	4.6907
1.9892	1.8546	1.3276
1.3140	2.2073	1.7435
3.4114	3.8632	1.5257
1.9102	4.4910	3.0311
4.4108	4.6255	1.7461
2.1316	5.4219	3.7365
4.7237	5.1998	1.7967
2.2738	6.2270	4.0297
5.2689	4.9217	2.1164
2.1931	6.4323	4.1315

Jika nilai kernel  $K$  dari data uji telah diperoleh, maka kita dapat mencari nilai  $f(\phi(x))$  yang dapat dilihat pada gambar 5.7.

Berikut merupakan langkah-langkah dalam pencarian nilai  $f(\phi(x))$ .

```
>> (W), (b), F =
[sum(Ku(1,:).*W)+b;sum(Ku(2,:).*W)+b;sum(
Ku(3,:).*W)+b;sum(Ku(4,:).*W)+b;sum(Ku(5,:).
.*W)+b;sum(Ku(6,:).*W)+b;sum(Ku(7,:).*W)+
b;sum(Ku(8,:).*W)+b;sum(Ku(9,:).*W)+b;sum(
Ku(10,:).*W)+b;sum(Ku(11,:).*W)+b;sum(Ku(12,:).
.*W)+b;sum(Ku(13,:).*W)+b;sum(Ku(14,:).
.*W)+b;sum(Ku(15,:).*W)+b;sum(Ku(16,:).*W)
)+b;sum(Ku(17,:).*W)+b;sum(Ku(18,:).*W)+b;
sum(Ku(19,:).*W)+b;sum(Ku(20,:).*W)+b;sum(
Ku(21,:).*W)+b;sum(Ku(22,:).*W)+b;sum(Ku(23,:).
.*W)+b;sum(Ku(24,:).*W)+b;sum(Ku(25,:).
.*W)+b;sum(Ku(26,:).*W)+b;sum(Ku(27,:).*W)
)+b;sum(Ku(28,:).*W)+b;sum(Ku(29,:).*W)+b;
sum(Ku(30,:).*W)+b;]
```



W = -3.8667 -0.9553 -1.6925 -0.1778  
-3.3555 -1.7816

b = 43.1227

F =

### Evaluasi

Setelah semua kalkulasi data selesai maka tergambar perbedaan antara label pada data uji dengan label hasil prediksi pada tabel 5.1.

**Tabel 4. Hasil Prediksi**

X1	X2	X3	Label Data Uji	Hasil Prediksi
0.85	0.42	0.9	1	-1
0.63	0.58	0.61	-1	-1
0.37	0.42	0.62	1	1
0.15	0.42	0.58	1	1
0.74	0.74	0.61	-1	-1
0.58	0.74	0.7	-1	-1
0.85	0.74	0.65	-1	-1
0.9	0.42	0.83	-1	-1
0.47	0.42	0.59	1	1
0.58	0.74	0.65	-1	-1
0.74	0.1	0.78	1	1
0.31	0.26	0.41	1	1
0.74	0.9	0.47	-1	-1
0.74	0.42	0.63	-1	-1
0.85	0.1	0.64	-1	1
0.31	0.26	0.59	1	1
0.85	0.42	0.7	-1	-1
0.9	0.42	0.9	1	-1
0.63	0.58	0.36	-1	1
0.15	0.58	0.62	1	1
0.37	0.58	0.4	-1	1
0.69	0.42	0.27	-1	1
0.74	0.58	0.65	-1	-1
0.85	0.58	0.61	-1	-1
0.85	0.74	0.83	-1	-1
0.26	0.26	0.17	-1	1
0.69	0.26	0.57	-1	1
0.63	0.42	0.82	1	-1
0.85	0.42	0.79	-1	-1
0.79	0.74	0.65	-1	-1

$$\text{Tingkat Akurasi} = \frac{13}{30} * 100\% = 43.33\%$$

Berdasarkan hasil perhitungan tingkat akurasi data uji yang diolah secara manual menggunakan metode *Support Vector Machine* (SVM) dengan *kernel trick polynomial* memiliki tingkat akurasi sebesar 43.33% di prediksi benar.

**Tabel 5. ....**

-14.9932	-4.8674	5.2180	10.9687	-12.2387
-10.5253	-16.5937	-14.2606	3.6638	-9.0517
0.0565	148737	-12.6073	-4.0371	1.0595
10.8843	-8.9535	-16.4409	1.8834	6.4034
6.8717	6.5818	-8.9239	-10.7461	-22.3797
20.6373	2.8495	-6.5360	-11.6289	-14.8685

### Simpulan

Kesimpulan yang diperoleh oleh penulis dari hasil penelitian tesis ini adalah :

1. Penerapan metode *Support Vector Machine* pada penelitian ini mampu memisahkan siswa yang berpotensi baik dan yang berpeluang *drop out*.
2. Metode ini mengolah kalkulasi data dari nilai akhir, perilaku, dan kehadiran. Dimana data tersebut sangat menentukan hasil prediksi siswa.
3. Hasil akhir dari penelitian ini menunjukkan bahwa jumlah siswa yang berpeluang *drop out* lebih banyak dibandingkan siswa yang mempunyai potensi.

### Saran

Berikut saran yang penulis berikan untuk untuk pengembangan tesis ini :

1. Variabel yang akan digunakan dalam menggunakan teknik *Support Vector Machine* harus jelas untuk ketepatan hasil kalkulasi data yang akan di peroleh.
2. Sangat diperlukan ketelitian dalam mengolah data menggunakan *Support Vector Machine* karena data yang diolah saling keterkaitan antar satu sama lainnya.
3. Penelitian ini mungkin mempunyai banyak kekurangan, dan penulis berharap penelitian ini dapat di kembangkan lagi. Sehingga kekurangan – kekurangan dalam penelitian yang ada menjadi lebih sempurna.

### DAFTAR PUSTAKA

- [1] Hastuti, Khafizh. (2012) “Teknik Klasifikasi Data Mining” Vol. 5, 33-41.

- [2] Sajadin Sembiring,. Ummul Khoir,. dan Lubis, Ihsan. (2011) *"Predictor Model in Prediction of Student Performance using Decision Tree and Smooth Support Vector Machine "* Sekolah Tinggi Harapan Medan.
- [3] Yulianton, Heribertus. (2008) *"Data Mining Untuk Dunia Bisnis"* Universitas Stikubank Semarang.
- [4] Widyarini, Tiananda., Budi Sentosa (2009) *"Aplikasi Metode Cross Entropy Untuk Support Vector Machine"*. Institut Teknologi Sepuluh Nopember (ITS) Surabaya.
- [5] Andari, Shofi., Santi W, Purnami., dan Bambang W, Otok (2013) *"Smooth Support Vector Machine dan Multivariate Adaptive Regression Spline untuk Mendiagnosis Kanker Payudara. Vol. 1, No 2.*
- [6] Tampubolon, Kennedi., Hoga, Saragih., dan Bobby, Reza. (2013) *"Implementasi Data Mining Algoritma Apriori pada Sistem Persediaan alat-alat Kesehatan"*. Vol. 1, No. 1.
- [7] Rachman, Farizi., dan Wulan Purnami, Santi. (2012) *"Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal dan Support Vector Machine"*, Institut Teknologi Sepuluh Nopember (ITS) Surabaya.
- [8] Wulan Purnami, Santi., Mohammad Zain, Jasni., dan Heriawan, Tutut. (2011) *"An Alternative Algorithm For Classification Large Categorical Dataset : k-mode Clustering Reduced Support Vector Machine"*. Universitas Malaysia Pahang.
- [9] Raharjo, Suwanto., dan Winarko, Edi. (2014) *"Klasterisasi, Klasifikasi dan Peringkasan Teks Berbahasa Indonesia"* Universitas Gunadarma.
- [10] Hartama, Dedy (2012) *"Model Aturan Keterhubungan Data Mahasiswa Dengan Algoritma Decision Tree"*, AMIK Tunas Bangsa Pematang Siantar.
- [11] Qin, Chuandong., dan Zhao, Huixia. (2014) *"Selecting The Optimal Combination Model Of FSSVM For The Imbalance Datasets"* Vol. 2014.
- [12] Huang, Chia-Hui., Keng-CHieh, Yang., dan Kao, Han-Ying. (2014) *"Analyzing Big Data With The Hybrid Interval Regression Methods"* Vol. 2014.
- [13] Sembiring, Sajadin., dkk. (2011) *"Prediction Of Student Academic Performance By An Application Of Data Mining Techniques"* Universitas Pahang Malaysia.