

**PENERAPAN METODE KLASIFIKASI *SUPPORT VECTOR MACHINE* (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG**

**Pusphita Anna Octaviani<sup>1</sup>, Yuciana Wilandari<sup>2</sup>, Dwi Ispriyanti<sup>3</sup>**

<sup>1</sup>Mahasiswa Jurusan Statistika FSM UNDIP

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM UNDIP

**ABSTRACT**

Accreditation is the recognition of an educational institution given by a competent authority, that is *Badan Akreditasi Nasional Sekolah/Madrasah* (BAN - S/M) after it is assessed that the institution has met the eight components of the accreditation assessment. An elementary school, as one of the compulsory basic education, should have the status of accreditation to ensure the quality of education. This study aimed to apply the classification method Support Vector Machine (SVM) on the data accreditation SD in Magelang. Support Vector Machine (SVM) is a method that can be used as a predictive classification by using the concept of searching hyperplane (separator functions) that can separate the data according to the class. SVM using the kernel trick for non-linear problems which can transform data into a high dimensional space using a kernel function, so that the data can be classified linearly. The results of this study indicate that the prediction accuracy of SVM classification using Gaussian kernel function RBF is 93.902%. It is calculated from 77 of 82 elementary schools that are classified correctly with the original classes.

**Keywords** : Accreditation, Classification, Support Vector Machine (SVM), hyperplane, Gaussian RBF Kernel, Accuracy

**1. PENDAHULUAN**

Dalam rangka penjaminan dan peningkatan mutu pendidikan nasional secara bertahap, terencana dan terukur sesuai amanat Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Pemerintah melakukan akreditasi untuk menilai kelayakan program dan/atau satuan pendidikan. Berkaitan dengan hal tersebut, Pemerintah telah menetapkan Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M) dengan Peraturan Mendiknas Nomor 29 Tahun 2005. Standar Nasional Pendidikan di Indonesia diatur dalam Peraturan Pemerintah (PP) No. 19 Tahun 2005. Di dalam PP tersebut, disebutkan bahwa lingkup Standar Nasional Pendidikan meliputi (1) Standar Isi, (2) Standar Proses, (3) Standar Kompetensi Lulusan, (4) Standar Pendidik dan Tenaga Kependidikan, (5) Standar Sarana dan Pra Sarana, (6) Standar Pengelolaan, (7) Standar Pembiayaan, dan (8) Standar Penilaian Pendidikan

Akreditasi menurut Kamus Besar Bahasa Indonesia adalah pengakuan terhadap lembaga pendidikan yang diberikan oleh badan yang berwenang setelah dinilai bahwa lembaga itu memenuhi syarat kebakuan atau kriteria tertentu. BAN-S/M sesuai dengan kewenangannya menerbitkan sertifikat akreditasi yang memuat nilai hasil akreditasi sebagai status akreditasi sekolah yang dinyatakan dalam huruf A (Amat baik), B (Baik), dan C (Cukup baik). Status akreditasi ini

berlaku untuk kurun waktu 5 tahun sejak tanggal ditetapkan. Setelah kurun waktu 5 tahun sekolah harus diakreditasi ulang.

Salah satu metode statistik yang dapat diterapkan untuk melakukan klasifikasi adalah *Support Vector Machine* (SVM). SVM merupakan suatu teknik untuk menemukan *hyperplane* yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). SVM memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan *support vector* sehingga proses komputasi menjadi cepat (Vapnik, 1995). Penelitian tentang SVM telah dilakukan oleh Rustam, *et al* (2003) yaitu membandingkan metode klasifikasi *K-Nearest Neighbor* (KNN) dengan metode SVM diperoleh kesimpulan bahwa SVM memiliki kinerja yang lebih unggul, karena telah mampu 100% mengklasifikasikan data aroma berdasarkan kelas yang tepat. Selain itu Rachman dan Purnami (2012) yang melakukan penelitian mengenai klasifikasi tingkat keganasan kanker dengan menggunakan metode regresi logistik dan SVM yang akhirnya diperoleh hasil bahwa tingkat akurasi menggunakan SVM lebih tinggi, yaitu sebesar 98,11%.

Pada penelitian ini penulis tertarik untuk menerapkan metode klasifikasi *Support Vector Machine* (SVM) pada akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. Tujuan dari penelitian ini adalah menerapkan metode klasifikasi *Support Vector Machine* pada data akreditasi Sekolah Dasar (SD) di Kabupaten Magelang.

## 2. TINJAUAN PUSTAKA

### 2.1 Standar Nasional Pendidikan

Standar Nasional Pendidikan adalah kriteria minimal tentang sistem pendidikan di seluruh wilayah hukum Negara Kesatuan Republik Indonesia. Pemerintah telah menetapkan Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M) dengan Peraturan Mendiknas Nomor 29 Tahun 2005. Akreditasi Menurut Kamus Besar Bahasa Indonesia adalah pengakuan terhadap lembaga pendidikan yang diberikan oleh badan yang berwenang setelah dinilai bahwa lembaga itu memenuhi syarat kebakuan atau kriteria tertentu. Berdasarkan penilaian delapan komponen standar pendidikan, maka sekolah akan mendapatkan sertifikat akreditasi dan status akreditasi. Status akreditasi tersebut berlaku selama lima tahun dan setelah itu akan dilakukan penilaian ulang untuk mendapatkan pengakuan akreditasi berdasarkan nilai yang baru.

### 2.2 Support Vector Machine

*Support Vector Machine* (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). SVM memiliki prinsip dasar linier *classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari *hyperplane* (*hyperplane*) yang dapat memaksimalkan jarak (margin) antara kelas data. Menurut Santosa (2007) *hyperplane* klasifikasi linier SVM dinotasikan :

$$f(x) = \mathbf{w}^T \mathbf{x} + b$$

sehingga menurut Vapnik dan Cortes (1995) diperoleh persamaan

$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1 \text{ untuk } y_i = +1$$

$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \leq -1 \text{ untuk } y_i = -1$$

dengan,  $\mathbf{x}_i$  = himpunan data training,  $i = 1, 2, \dots, n$  dan  $y_i$  = label kelas dari  $\mathbf{x}_i$   
 Untuk mendapatkan *hyperplane* terbaik adalah dengan mencari *hyperplane* yang terletak di tengah-tengah antara dua bidang pembatas kelas dan untuk mendapatkan *hyperplane* terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda (Santosa, 2007). Margin dapat dihitung dengan  $\frac{2}{\|\mathbf{w}\|}$

Mencari *hyperplane* terbaik dapat digunakan metode *Quadratic Programming* (QP) *Problem* yaitu meminimalkan

$$\frac{1}{2} \mathbf{w}^T \mathbf{w}$$

dengan syarat  $y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, 3, \dots, n$

Solusi untuk mengoptimasi oleh Vapnik (1995) diselesaikan dengan menggunakan fungsi *Lagrange* sebagai berikut

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i \{y_i [(\mathbf{w}^T \cdot \mathbf{x}_i) + b] - 1\} \quad (1)$$

dengan  $\alpha_i$  = pengganda fungsi *Lagrange* dan  $i = 1, 2, \dots, n$

Nilai optimal dapat dihitung dengan memaksimalkan  $L$  terhadap  $\alpha_i$ , dan meminimalkan  $L$  terhadap  $\mathbf{w}$  dan  $b$ . Hal ini seperti kasus dual problem

$$\max_{\alpha} W(\alpha) = \max_{\alpha} (\min_{\mathbf{w}, b} L(\mathbf{w}, b, \alpha)) \quad (\text{Gunn, 1998}).$$

Nilai minimum dari fungsi lagrange tersebut diberikan oleh

$$\begin{aligned} \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L}{\partial \mathbf{w}} = 0 &\Rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i \end{aligned} \quad (2)$$

Untuk menyederhanakannya persamaan (1) harus ditransformasikan ke dalam fungsi *Lagrange Multiplier* itu sendiri, sehingga menurut Santosa (2007) persamaan (1) menjadi

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i) - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \quad (3)$$

Berdasarkan persamaan (2), maka persamaan (3) oleh Hastie *et al* (2001) menjadi sebagai berikut

$$L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

dan diperoleh dual problem

$$\max_{\alpha} L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

dengan batasan,  $\alpha_i \geq 0, \quad i = 1, 2, \dots, n$  dan  $\sum_{i=1}^n \alpha_i y_i = 0$

Data training dengan  $\alpha_i > 0$  terletak pada *hyperplane* disebut *support vector*. Data training yang tidak terletak pada *hyperplane* tersebut mempunyai  $\alpha_i = 0$ . Setelah solusi permasalahan *quadratic programming* ditemukan (nilai  $\alpha_i$ ), maka kelas dari data yang akan diprediksi atau data testing dapat ditentukan berdasarkan nilai fungsi berikut.

$$f(\mathbf{x}_t) = \sum_{s=1}^{ns} \alpha_s y_s \mathbf{x}_s \cdot \mathbf{x}_t + b$$

dengan

- $\mathbf{x}_t$  = data yang akan diprediksi kelasnya (data testing)
- $\mathbf{x}_s$  = data *support vector*,  $s = 1, 2, \dots, ns$
- ns = banyak data *support vector*

Pada kasus linier *non-separable* beberapa data mungkin tidak bisa dikelompokkan secara benar atau terjadi *misclassification* (Novianti dan Purnami, 2012). Sehingga persamaan dimodifikasi dengan menambahkan variabel *slack*

$\xi_i \geq 0$ . Variabel *slack* ini merupakan sebuah ukuran kesalahan klasifikasi (Rachman dan Purnami, 2012). Berikut ini adalah pembatas yang sudah dimodifikasi oleh Gunn (1998) untuk kasus *non-separable*:

$$y_i[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, n$$

Sehingga persamaan (2.6) menjadi meminimalkan

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i$$

dengan syarat persamaan (2.12) dan  $\xi_i \geq 0, i = 1, 2, \dots, n$

Banyak teknik data mining atau *machine learning* yang dikembangkan dengan asumsi kelinieran, sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier (Santosa, 2007). SVM dapat bekerja pada data non-linier dengan menggunakan pendekatan kernel pada fitur data awal himpunan data. Fungsi kernel yang digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang relatif lebih tinggi).

Menurut Prasetyo (2012) macam fungsi kernel diantaranya:

1. Kernel *Gaussian Radial Basic Function* (RBF)

$$\mathbf{K}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

2. Kernel *Polynomial*

$$\mathbf{K}(x_i, x_j) = \left((x_i \cdot x_j) + c\right)^d$$

$x_i$  dan  $x_j$  adalah pasangan dua data training. Parameter  $\sigma, c, d > 0$  merupakan konstanta. Fungsi kernel mana yang harus digunakan untuk substitusi *dot product* di *feature space* sangat tergantung pada data karena fungsi kernel ini akan menentukan fitur baru di mana *hyperplane* akan dicari (Santosa, 2007).

Pada awalnya SVM dikembangkan untuk persoalan klasifikasi dua kelas, kemudian dikembangkan kembali untuk klasifikasi multikelas (Santosa, 2007). Dalam klasifikasi kasus multikelas, *hyperplane* yang terbentuk adalah lebih dari satu. Salah satu metode pendekatan yang digunakan adalah satu lawan semua (SLA). Metode SLA untuk kasus klasifikasi  $k$ -kelas, menemukan  $k$  *hyperplane* dimana  $k$  adalah banyak kelas dan  $\rho$  adalah *hyperplane*. Dalam metode ini  $\rho^{(\ell)}$  diujikan dengan semua data dari kelas  $\ell$  dengan label +1, dan semua data dari kelas lain dengan label -1.

Konsep pada SLA yaitu dimisalkan pada kasus tiga kelas, kelas 1, 2, dan 3. Bila akan diujikan  $\rho^{(1)}$ , semua data dalam kelas 1 diberi label +1 dan data dari kelas 2 dan 3 diberi label -1. Pada  $\rho^{(2)}$ , semua data dalam kelas 2 diberi label +1 dan data dari kelas 1 dan 3 diberi label -1. Begitu juga untuk  $\rho^{(3)}$ , semua data dalam kelas 3 diberi label +1 dan data dari kelas 1 dan 2 diberi label -1. Kemudian dicari *hyperplane* dengan algoritma SVM dua kelas. Maka akan didapat *hyperplane* untuk masing-masing kelas diatas. Kemudian kelas dari suatu data baru  $x$  menurut Hsu dan Lin (2002) ditentukan berdasarkan nilai terbesar dari *hyperplane*

$$\text{kelas } x = \arg \max_{\ell=1, \dots, k} \left( (\mathbf{w}^{(\ell)})^T \cdot \Phi(x) + b^{(\ell)} \right)$$

### 2.3 Pengukuran Kinerja Klasifikasi

Pengukuran kinerja klasifikasi pada data asli dan data hasil dari model klasifikasi dilakukan dengan menggunakan tabulasi silang (*matriks konfusi*) yang berisi informasi tentang kelas data asli yang direpresentasikan pada baris matriks dan kelas data hasil prediksi suatu algoritma direpresentasikan pada kolom klasifikasi (Prasetyo, 2012). Berikut merupakan contoh matriks konfusi

**Tabel 1.** Matriks Konfusi Tiga Kelas

$F_{gh}$		Kelas Prediksi (h)		
		Kelas 1	Kelas 2	Kelas 3
Kelas Asli (g)	Kelas 1	$F_{11}$	$F_{12}$	$F_{13}$
	Kelas 2	$F_{21}$	$F_{22}$	$F_{23}$
	Kelas 3	$F_{31}$	$F_{32}$	$F_{33}$

Menurut Prasetyo (2012) ketepatan klasifikasi dapat dilihat dari akurasi klasifikasi. Akurasi klasifikasi menunjukkan performansi model klasifikasi secara keseluruhan, dimana semakin tinggi akurasi klasifikasi hal ini berarti semakin baik performansi model klasifikasi.

$$\begin{aligned} \text{Akurasi klasifikasi} &= \frac{F_{11}+F_{22}+F_{33}}{F_{12}+F_{13}+F_{21}+F_{23}+F_{31}+F_{32}+F_{11}+F_{22}+F_{33}} \\ \text{Error klasifikasi} &= \frac{F_{12}+F_{13}+F_{21}+F_{23}+F_{31}+F_{32}}{F_{12}+F_{13}+F_{21}+F_{23}+F_{31}+F_{32}+F_{11}+F_{22}+F_{33}} \end{aligned}$$

Semua klasifikasi berusaha untuk membentuk model dengan nilai akurasi tinggi dan laju error rendah.

## 3. METODOLOGI PENELITIAN

### 3.1 Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data tentang nilai akreditasi Sekolah Dasar (SD) di Kabupaten Magelang mulai tahun penetapan 2011-2013 yang diperoleh dari website resmi BAN-S/M dengan status akreditasi A, B dan C. Untuk variabel dependen (Y) terdiri dari tiga kategori, yaitu :

- Y = 1, untuk sekolah yang memiliki akreditasi A
- Y = 2, untuk sekolah yang memiliki akreditasi B
- Y = 3, untuk sekolah yang memiliki akreditasi C

Pada penelitian ini, hanya digunakan tiga kategori tersebut diatas dikarenakan di Kabupaten Magelang tidak ada Sekolah Dasar (SD) yang berstatus akreditasi TT (Tidak Terakreditasi). Sedangkan untuk variabel independen terdiri dari 8 komponen standar yaitu Standar Isi ( $X_1$ ), Standar Proses ( $X_2$ ), Standar Kompetensi Lulusan ( $X_3$ ), Standar Pendidik dan Tenaga Kependidikan ( $X_4$ ), Standar Sarana dan Prasarana ( $X_5$ ), Standar Pengelolaan ( $X_6$ ), dan Standar penilaian pendidikan ( $X_8$ ).

### 3.2 Langkah-langkah Analisis

Langkah-langkah analisis adalah sebagai berikut

1. Melakukan pengumpulan data sekunder nilai akreditasi Sekolah Dasar (SD) di Kabupaten Magelang mulai tahun 2011-2013 dengan status akreditasi A, B, dan C.
2. Melakukan prediksi klasifikasi status akreditasi sekolah dasar (SD) di Kabupaten Magelang menggunakan metode analisis SVM dengan variabel

dependen dan independen yang telah ditentukan. Berikut ini langkah-langkah penyelesaiannya

- a. Memasukkan data sesuai format software *MATLAB*.
  - b. Membagi data menjadi data training dan data testing, yaitu data tahun 2011-2012 sebagai data training dan data tahun 2013 sebagai data testing.
  - c. Menentukan metode pendekatan *hyperplane* SVM Multikelas yaitu dengan metode SVM Multikelas satu lawan semua (SLA)
  - d. Menentukan fungsi kernel yang akan digunakan sebagai permodelan *hyperplane* SVM SLA.
  - e. Menentukan nilai parameter C dan nilai-nilai parameter kernel yang akan digunakan sebagai permodelan *hyperplane* SVM SLA.
  - f. Mendapatkan nilai alpha dan b
  - g. Membentuk tiga persamaan *hyperplane*
  - h. Melakukan prediksi klasifikasi
3. Evaluasi performansi model klasifikasi menggunakan matriks konfusi.
    - a. Menghitung akurasi klasifikasi hasil prediksi
    - b. Memilih nilai parameter dan fungsi kernel terbaik

#### 4. PEMBAHASAN

##### 4.1 Analisis Deskriptif Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data tentang nilai status akreditasi Sekolah Dasar (SD) di Kabupaten Magelang mulai tahun penetapan akreditasi 2011 - 2013 yaitu sebanyak 419 sekolah dengan status akreditasi A,B, dan C. Data diambil dari website resmi BAN-S/M pada bulan Februari 2014.

Pada tahun 2013 sebanyak 16 SD berstatus akreditasi A, 66 SD berstatus akreditasi B dan tidak ada SD yang berakreditasi C. Tahun 2012 sebanyak 41 SD berakreditasi A, 153 SD berstatus akreditasi B, dan hanya 1 SD yang berstatus akreditasi C. Sedangkan untuk tahun 2011, 24 SD berstatus akreditasi A, 114 SD berstatus akreditasi B dan 4 SD berstatus akreditasi C. Dengan demikian sebanyak 19.33% berakreditasi A, 79.47% berakreditasi B dan 1.19% berakreditasi C.

##### 4.2 Pembentukan Persamaan SVM Multikelas Satu Lawan Semua (SLA)

Pada kasus SVM Multikelas SLA jumlah persamaan yang terbentuk yaitu sebanyak  $k$  atau jumlah kelas. Jadi untuk kasus tiga kelas, maka akan terbentuk tiga persamaan, yaitu  $f_1(x)$ ,  $f_2(x)$ , dan  $f_3(x)$ .  $f_1(x)$  adalah persamaan untuk data training dalam kelas 1 diberi label +1, dan data dalam kelas 2 dan 3 diberi label -1.  $f_2(x)$  adalah persamaan untuk data training dalam kelas 2 diberi label +1, dan data dalam kelas 1 dan 3 diberi label -1.  $f_3(x)$  adalah persamaan untuk data training dalam kelas 3 diberi label +1, dan data dalam kelas 1 dan 2 diberi label -1. Sehingga dalam setiap pembentukan persamaan, kelas yang awalnya ada 3 kelas menjadi 2 kelas.

Langkah pembentukan persamaan SVM Multikelas *Gaussian* RBF dibantu dengan *software MATLAB 2009a* adalah sebagai berikut

1. Membagi keseluruhan data yang berjumlah 419 data menjadi data training dan data testing. Data training yang akan digunakan sebagai permodelan awal sebanyak 337 data dan data testing untuk prediksi sebanyak 82 data.
2. Persamaan pertama yang akan dibentuk yaitu  $f_1(x)$ , sehingga data training dalam kelas 1 diberi label +1, dan data dalam kelas 2 dan 3 diberi label -1.

- Pemilihan metode SVM yang digunakan adalah SVM Multikelas SLA dan fungsi kernel yang digunakan adalah *Gaussian RBF* dengan parameter  $\sigma = 2$  dan  $C = 1$  dan *Polynomial* adalah  $d = 2, C = 1$
- Setelah ditentukan fungsi kernel dan parameternya, dengan bantuan software maka parameter itu digunakan untuk memetakan data training menggunakan fungsi kernel *Gaussian RBF* berikut.

$$\mathbf{K}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

atau fungsi kernel *Polynomial* berikut

$$\mathbf{K}(x, x_i) = (1 + x^T \cdot x_i)^2$$

- Setelah data training dipetakan menggunakan fungsi kernel, kemudian mendapatkan nilai  $\alpha$  dan  $b$  melalui *Quadratic Programming (QP)* dengan bantuan software *MATLAB 2009a*. Nilai  $\alpha$  dan  $b$  selanjutnya digunakan dalam persamaan SVM untuk memprediksi klasifikasi data testing.

Persamaan untuk SVM Kernel *Gaussian RBF* adalah

$$f(x) = \mathbf{w}^T \cdot \mathbf{x} + b$$

dimana  $\mathbf{w}^T \cdot \mathbf{x} + b = \alpha^T \mathbf{y}^T \mathbf{K}(x_{training}, x_{testing}) + b$

untuk persamaan  $f_1(x)$  adalah sebagai berikut

$$\begin{aligned} f_1(x) &= \alpha^T \mathbf{y}^T \mathbf{K}(x_{training}, x_{testing}) + b \\ &= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{337 \times 1}^T \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{bmatrix}_{337 \times 337} \mathbf{K}(x_{training}, x_{testing})_{337 \times 82} + 0 \end{aligned}$$

$$\begin{aligned} \text{dengan } \mathbf{K}(x_{training}, x_{testing}) &= \exp\left(-\frac{\|x_{training} - x_{testing}\|^2}{2(2)^2}\right) \\ &= \exp\left(-\frac{\|x_{training} - x_{testing}\|^2}{8}\right) \end{aligned}$$

Sedangkan persamaan untuk SVM Kernel *Polynomial* adalah

$$f(x) = \mathbf{w}^T \cdot \mathbf{x} + b$$

$$f_1(x) = (\sum_{i=1}^{36} \alpha_i y_i \Phi(x_i) \cdot \Phi(x) + b), \text{ dengan } b = 0$$

$$\begin{aligned} &= (0.0001 \times 1 \times \Phi(x_1) \cdot \Phi(x)) + (0.0029 \times 1 \times \Phi(x_2) \cdot \Phi(x)) + \dots \\ &\quad + (0.0043 \times -1 \times \Phi(x_{36}) \cdot \Phi(x)) \end{aligned}$$

Dari perhitungan menggunakan persamaan  $f_1(x)$  didapatkan vektor baris dengan ukuran  $1 \times 82$  (sebanyak data testing) yang merupakan prediksi klasifikasi sementara. Selanjutnya membentuk persamaan kedua  $f_2(x)$  dan  $f_3(x)$  dengan langkah yang sama seperti dalam persamaan  $f_1(x)$ .

### 4.3 Analisis Klasifikasi Akreditasi Sekolah Dasar (SD) di Kab. Magelang

Klasifikasi SVM Multikelas SLA dengan fungsi kernel *Gaussian RBF* menggunakan 5 nilai parameter  $\sigma$  sebagai analisis  $\sigma = 1, \sigma = 2, \sigma = 3, \sigma = 4, \sigma = 5$  dengan nilai parameter  $C = 1, C = 5, C = 10, C = 50, C = 100$ .

Hasil akurasi klasifikasi kernel *Gaussian RBF* pada data training dapat dilihat pada Tabel 2. Terlihat bahwa model yang dibentuk data training dapat digunakan untuk memprediksi kelas data training itu sendiri dengan akurasi sebesar 100% yang berarti kelas dapat diklasifikasikan tepat sesuai dengan kelas aslinya dengan error sebesar 0%. Berikut adalah tabel hasil akurasi untuk prediksi kelas pada data training

**Tabel 2.** Hasil Akurasi Klasifikasi Kernel *Gaussian* RBF pada Data Training

C	RBF					
	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$	$\sigma=6$
1	100	100	100	100	100	100
5	100	100	100	100	100	100
10	100	100	100	100	100	100
50	100	100	100	100	100	100
100	100	100	100	100	100	100

Selanjutnya untuk hasil akurasi klasifikasi kelas menggunakan data testing adalah sebagai berikut

**Tabel 3.** Hasil Akurasi Klasifikasi SVM Kernel *Gaussian* RBF pada Data Testing

C	RBF				
	$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$
1	91.463	92.683	93.902	92.683	93.902
5	91.463	92.683	93.902	93.902	93.902
10	91.463	92.683	93.902	93.902	93.902
50	91.463	92.683	93.902	93.902	93.902
100	91.463	92.683	93.902	93.902	93.902

Tampak bahwa akurasi mencapai nilai maksimal saat  $\sigma = 3$ , sehingga prediksi terbaik yaitu dipilih nilai parameter  $\sigma = 3$  untuk nilai  $C = 1$ , yaitu nilai  $C$  paling kecil dengan akurasi sebesar 93.902 %.

Berikut adalah matriks konfusi Kernel *Gaussian* RBF  $\sigma = 3$ ,  $C = 1$  pada data testing

**Tabel 4.** Matriks Konfusi Kernel *Gaussian* RBF  $\sigma = 3$ ,  $C = 1$  pada Data Testing

$F_{gh}$		Kelas Hasil Prediksi ( $h$ )		
		<i>Kelas 1</i>	<i>Kelas 2</i>	<i>Kelas 3</i>
Kelas Asli ( $g$ )	<i>Kelas 1</i>	12	4	0
	<i>Kelas 2</i>	1	65	0
	<i>Kelas 3</i>	0	0	0

Berdasarkan Tabel 4. diperoleh akurasi klasifikasi sebesar 93.902 % dan error klasifikasi sebesar 6.098 %. Akurasi 93.902 % menunjukkan bahwa dapat mengklasifikasikan 77 SD dari 82 SD secara benar sama dengan kelas aslinya. Sedangkan 4 SD lainnya diklasifikasikan berbeda dengan data asli atau error klasifikasi sebesar 6.098 %.

Pada klasifikasi ini dicobakan menggunakan fungsi kernel Polynomial dengan menggunakan dua nilai  $d$  pada fungsi Kernel Polynomial yaitu  $d = 1$  dan  $d = 2$  dengan nilai parameter  $C = 1$ ,  $C = 5$ ,  $C = 10$ ,  $C = 50$ ,  $C = 100$  dan akan dibandingkan ketepatan klasifikasi untuk masing-masing parameter fungsi kernel polynomial. Berikut adalah tabel hasil akurasi untuk prediksi kelas pada data training.



**Tabel 5.** Hasil Akurasi Klasifikasi SVM Kernel *Polynomial* pada Data Training

C	Polynomial	
	d=1	d=2
1	79.228	98.220
5	89.021	98.810
10	93.472	98.810
50	96.736	98.810
100	96.736	98.810

Terlihat pada Tabel 5. bahwa akurasi terbaik untuk pengklasifikasian data training adalah sebesar 98.810 % untuk parameter  $d = 2$  dan  $C = 5$ . Berikut hasil akurasi klasifikasi pada data testing menggunakan fungsi kernel *polynomial*.

**Tabel 6.** Hasil Akurasi Klasifikasi SVM Kernel *Polynomial* pada Data Testing

C	Polynomial	
	d=1	d=2
1	80.488	85.366
5	87.805	84.146
10	91.463	78.049
50	92.683	84.146
100	92.683	84.146

Berdasarkan Tabel 6. terlihat bahwa klasifikasi terbaik menggunakan fungsi kernel *polynomial* adalah memakai  $d = 1$  dan  $C = 50$  dengan tingkat akurasi klasifikasi terbesar. Untuk mengetahui jumlah kelas yang diprediksi secara benar atau secara salah, dapat dilihat dari matriks konfusi berikut ini

**Tabel 7.** Matriks Konfusi Kernel *Polynomial*  $d = 1$ ,  $C = 50$  pada Data Testing

$F_{gh}$		Kelas Hasil Prediksi ( $h$ )		
		<i>Kelas 1</i>	<i>Kelas 2</i>	<i>Kelas 3</i>
Kelas Asli ( $g$ )	<i>Kelas 1</i>	13	3	0
	<i>Kelas 2</i>	3	63	0
	<i>Kelas 3</i>	0	0	0

Berdasarkan Tabel 7. diperoleh akurasi klasifikasi sebesar 92.682 % dan error klasifikasi sebesar 7.317 %. Artinya dengan akurasi 92.682 % sebanyak 76 SD dari 82 SD dapat diklasifikasikan secara benar sesuai dengan kelas asli, sedangkan sisanya sebanyak 6 SD diklasifikasikan berbeda dengan kelas asli.

## 5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan yaitu

1. Klasifikasi Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang menggunakan metode *Support Vector Machine* (SVM) dengan data training yang diujikan sebanyak 337 data memiliki akurasi klasifikasi sebesar 100% menggunakan fungsi kernel *Gaussian Radial Basic*

- Function* (RBF). Sedangkan menggunakan fungsi kernel *Polynomial* akurasi klasifikasi adalah sebesar 98.810 %.
2. Pada pengujian dengan data testing sebanyak 82 data, akurasi klasifikasi yang didapat yaitu sebesar 93.902% menggunakan kernel *Gaussian Radial Basic Function* (RBF). Sedangkan menggunakan fungsi kernel *Polynomial* akurasi klasifikasinya adalah sebesar 92.683 % .
  3. Dengan demikian akurasi klasifikasi terbaik yaitu menggunakan fungsi kernel *Gaussian Radial Basic Function* (RBF) karena menghasilkan akurasi yang lebih besar, yaitu dapat mengklasifikasikan 337 SD dari 337 SD secara benar sesuai dengan kelas asli pada data training. Sedangkan pada data testing dapat mengklasifikasikan 77 SD dari 82 SD secara benar sesuai dengan kelas asli.

#### DAFTAR PUSTAKA

- Gunn, S. R. 1998. *Support Vector Machines for Classification and Regression*. Southampton: University of Southampton.
- Hastie, T., Tibshirani, R., dan Friedman, J. 2001. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. California : Springer
- Hsu, C.W. dan Lin, C.J. 2002. A Comparison of Methods for Multi-class Support Vector Machines. *IEEE Transaction on Neural Network*, 13(2) : 415-425.
- Novianti, F.A dan Purnami, S.W. 2012. Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan *Support Vector Machine* (SVM) Berdasarkan Hasil Mamografi. *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1 ISSN : 2301-928X.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Andi : Yogyakarta
- Rachman, F dan Purnami, S.W. 2012. Klasifikasi Tingkat Keganasan *Breast Cancer* dengan Menggunakan Regresi Logistik Ordinal dan *Support Vector Machine* (SVM). *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1 ISSN : 2301-928X.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu : Yogyakarta.
- Vapnik, V dan Cortes, C. 1995. Support Vector Networks. *Machine Learning*, 20, 273-297.
- [BAN – S/M]. 2014. Badan Akreditasi Nasional Sekolah/Madrasah. <http://ban-sm.or.id>. (diakses tanggal 5 Februari 2014).