**KLASIFIKASI DATA SKALA BESAR DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM) MENGGUNAKAN *KERNEL* BERBASIS *GAUSSIAN***

**Irfan Juni1, Armin Lawi2, Sri Astuti Thamrin3**

**1,2,3**Jurusan Matematika , Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin

**ABSTRAK**

Pengolahan data skala besar kemudian menggunakan teknik *data mining*. *Support Vector Machine* (*SVM*) adalah salah satu metode klasifikasi dalam *data mining*. Ide dasar dari metode *SVM* adalah menemukan bidang pemisah linier yang membagi data ke dalam dua kelas. Kendala yang terjadi adalah seringkali sekumpulan data sulit untuk diklasifikasi ke dalam dua kelas dengan bidang pemisah linier. Pada penelitian ini, untuk mengatasi kendala aplikasi metode SVM pada data yang tidak dapat dipisah secara linier digunakan metode *kernel* berbasis *Gaussian*. Metode *kernel* lain seperti *kernel* linier dan polinomial juga akan dibahas untuk dibandingkan dengan metode *kernel Gaussian*. Hasil dari penelitian ini menunjukan bahwa penggunaan kernel Gaussian dapat meningkatkan tingkat akurasi dari klasifikasi data menggunakan SVM.

*Kata kunci: Data mining, SVM, metode kernel, kernel Gaussian, tingkat akurasi, kernel linier, kernel polinomial*

1. **Pendahuluan**

Kehidupan modern saat ini banyak sekali berhubungan dengan manajemen data dalam skala besar. Data skala besar adalah sekumpulan data yang sangat besar dan beragam yang dalam pengolahannya akan menemui kesulitan jika menggunakan metode atau aplikasi standar statistika (Snijders, 2002). Pengolahan data skala besar kemudian dilakukan dengan teknik *data mining.*

Salah satu metode penggalian informasi dalam teknik *data mining* adalah klasifikasi (Charu, 1999). *Support Vector Machine* (*SVM*) termasuk dalam metode klasifikasi data mining. *SVM* diperkenalkan pertama kali pada tahun 1979 oleh Vapnik (Boswell, 2002). Setelah tahun 1992, SVM berkembang dengan pesat. SVM adalah salah satu teknik yang relatif baru dibandingkan dengan teknik lain, tetapi memiliki performa yang lebih baik di berbagai bidang aplikasi seperti bioinformatika, pengenalan tulisan tangan, klasifikasi teks dan lain sebagainya (Christianini dan Taylor, 2000). Aplikasi SVM pada data skala besar seringkali menemui kendala dimana terdapat data yang sulit atau bahkan tidak memungkinkan untuk dipisahkan secara linier. Solusi dari permasalahan ini adalah dengan menambahkan metode *kernel* pada aplikasi *SVM* (Santoso, 2012). *Kernel Gaussian* adalah salah satu kernel yang sering digunakan dalam klasifikasi data *SVM*. Hal ini disebabkan karena *kernel Gaussian* memiliki tingkat akurasi yang paling besar dibanding kernel lain seperti polinomial atau linier (Borivikov, 1999).

Ide dasar dari metode SVM adalah untuk menemukan bidang pemisah terbaik yang dapat memisahkan data ke dalam dua buah kelas dalam sebuah ruang fitur (feature space) (Nugroho, 2008).

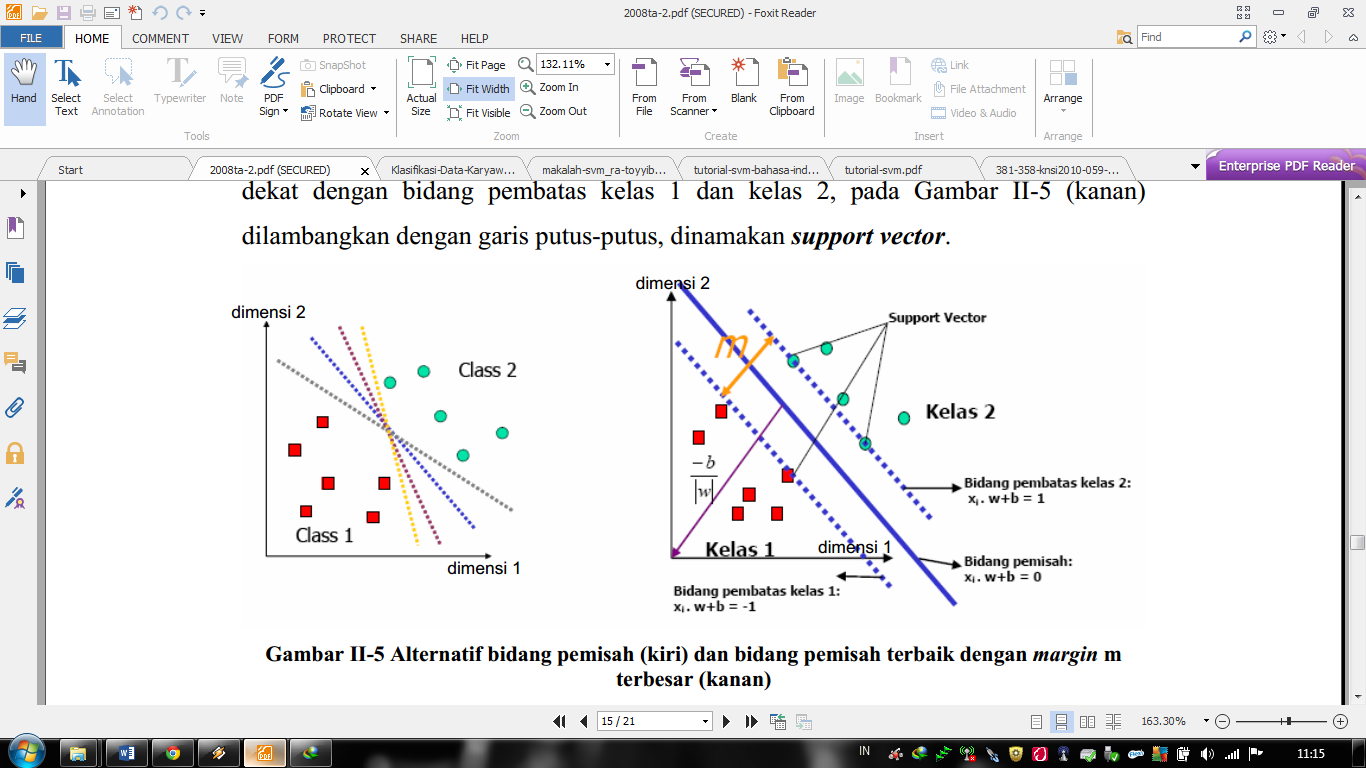
Penelitian ini bertujuan untuk mengklasifikasi data skala besar dengan metode *SVM.* Selain itu, penelitian ini juga untuk mengetahui tingkat akurasi klasifikasi data dengan *SVM* berbasis *kernel Gaussian* dan membandingkannya dengan *kernel* lain seperti polinomial atau linier.

**2. Tinjauan Pustaka**

* 1. ***Support Vector Machine* (*SVM*)**

*SVM* adalah salah satu dari beberapa teknik klasifikasi dalam data mining. Teori yang mendasari SVM sendiri telah berkembang sejak tahun 1960-an, tetapi baru diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992.

Gambar 1.(a) memperlihatkan berbagai alternatif bidang pemisah yang dapat dibuat untuk memisahkan data ke dalam dua buah kelas. Data yang paling dekat dengan bidang pemisah kelas 1 dan kelas 2 pada Gambar 1.(b) dilambangkan dengan garis putus-putus disebut *support vector* (Nugroho, 2008).



(a) (b)

**Gambar 1.** Alternatif bidang pemisah (a) dan bidang pemisah terbaik (b) dengan margin (*m*) terbesar.

(Sumber: Fadli, 2008, hal 15)

Menurut Burges (1998) bidang pembatas untuk setiap kelas memenuhi persamaan:

= +1

= -1. (2.1)

Sedangkan menurut Gunn (1998) margin(*m*) atau jarak dari kedua bidang pembatas pada persamaan (2.1) dapat dituliskan seperti berikut:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Bidang pemisah terbaik adalah bidang pemisah yang memiliki margin (*m*) yang paling besar. Oleh sebab itu margin (*m*) pada persamaan (2.2) harus dimaksimumkan namun tetap memperhatikan persamaan (2.1). Menurut Abe (2006) persamaan (2.1) dan (2.2) dapat digeneraliasasi menjadi masalah optimasi dengan fungsi objektif sebagai berikut:

|  |  |  |
| --- | --- | --- |
| min  dengan kendala , untuk setiap *i=1,...,m.* |  |  |

Persamaan (2.3) akan lebih mudah jika dibawa kedalam fungsi *Lagrange*. Fungsi *Lagrange* dibentuk dengan mengalikan fungsi kendala dengan pengali *Lagrange* ( kemudian dikurangkan dengan fungsi asal (Hofmann, 2006). Bentuk fungsi *Lagrange* dari persamaan (2.3) adalah:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Persamaan (2.4) sulit untuk diketahui solusinya sebab masih terdapat elemen *w* dan *b* yang belum diketahui. Hal ini menyebabkan persamaan (2.4) harus dibawa ke dalam bentuk dual.

Bentuk dual dari persamaan (2.4) adalah:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Persamaan (2.5) digunakan dalam proses pelatihan untuk mendapatkan nilai . Setelah mendapatkan nilai , maka akan ditemukan nilai *w*. Setelah mendapatkan nilai *w* maka nilai *b* (Boswell, 2002) adalah:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Karena nilai *b* dan *w* telah diketahui maka fungsi pemisah (*f*(*x*)) untuk menentukan kelas dari suatu data *x* (Abe, 2006) adalah:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

* 1. **Metode *Kernel***

Metode *kernel* adalah solusi yang digunakan dalam aplikasi *SVM* ketika data yang akan diklasifikasi sulit atau bahkan tidak mungkin untuk diklasifikasikan dengan bidang pembatas linier. Metode kernel untuk *SVM* diperkenalkan tahun 1992 oleh Boser, Guyon, dan Vapnik (Christianini, 2002). Penggunaan metode *kernel* menyebabkan suatu data dalam ruang *input* dipetakan ke ruang fitur *F* dengan dimensi yang lebih tinggi melalui map sebagai berikut: (Santoso, 2012, hal 8). Pemetaan ini dilakukan dengan tetap menjaga karakteristik atau topologi data.

Kendala yang terjadi adalah seringkali fungsi tidak tersedia atau sulit untuk dihitung sebab data akan dipetakan ke dimensi yang lebih tinggi sampai data ini dapat dipisahkan secara linier. Permasalahan ini tidak berpengaruh pada perhitungan fungsi pemisah untuk *SVM*. Hal ini karena berdasarkan persamaan (2.5), perhitungan fungsi pemisah terbaik melibatkan *dot product* dari 2 buah vektor. Menurut Santoso (2012) *dot product* dari dua buah vektor dalam metode *SVM* dapat dihitung baik dalam ruang *input* maupun ruang fitur. Hal ini berarti meskipun tidak diketahui namun *dot product* tetap dapat dihitung (Santoso, 2012*,* hal9).

Fungsi kernel didefinisikan sebagai :

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Menurut Santoso (2012) *dot product* dari fungsi kernel di atas adalah

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Setelah data dipetakan ke ruang fitur, fungsi pemisah juga akan berubah menurut hasil pemetaan ke ruang fitur tersebut. Berdasarkan persamaan (2.5) elemen terakhir yaitu adalah *dot product*, dalam hal ini sesuai dengan persamaan (2.7), sehingga persamaan (2.5) menjadi:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Beberapa bentuk kernel yang umum digunakan pada metode *SVM* adalah:

1. Linier K(
2. Polinomial: K(
3. *Gaussian* : K(
4. Sigmoid: K(
   1. ***Kernel Gaussian***

Menurut Borovikov (1999), diantara beberapa *kernel* *SVM* yang sering digunakan yaitu linier, polinomial dan *Gaussian*, *kernel* *Gaussian* adalah *kernel* yang mempunyai tingkat keakuratan yang paling baik.

Fungsi *kernel* *Gaussian* (Albatal dan Litle, 2013) adalah:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

dimana adalah variansi. Seringkali dituliskan sebagai variabel seperti berikut ini (Hsu *et al.*, 2010):

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

sehingga persamaan (2.10) menjadi (Hsu *et al*., 2010):

|  |  |  |  |
| --- | --- | --- | --- |
|  | *,* |  |  |

**3. Hasil dan Pembahasan**

* 1. **Sumber dan jenis data**

Data yang digunakan pada penelitian ini terdiri dari dua buah *dataset* dengan rincian sebagai berikut:

*Dataset* 1 berupa data tentang tinggi dan berat badan dari pemain basket dan pemain football di Amerika. Data pemain basket diunduh dari situs [*http://www.sportsillustrated.cnn.com/basketball/nba/teams/*](http://www.sportsillustrated.cnn.com/basketball/nba/teams/) sedangkan data pemain football diunduh dari situs *http://chicagosports.sportsdirectinc.com /football/nfl-teams.aspx? page=/data/nfl/teams/rosters/roster2.html*

*Dataset* 2 berupa data tentang batasan-batasan dalam urutan DNA yang diunduh dari situs [http://www.csie.ntu.edu.tw/~cjlin/lib*SVM*tools/*dataset*s/binary.html# fourclass](http://www.csie.ntu.edu.tw/~cjlin/libSVMtools/datasets/binary.html# fourclass)

Atribut untuk *dataset* 1 berjumlah dua buah yaitu tinggi dan berat badan. Tinggi badan pemain basket dan *football* diukur dengan satuan sentimeter (cm) sedangkan berat badan diukur dengan kilogram (kg). *Dataset* 2 merupakan batas-batas dalam urutan DNA. Batas-batas ini dibedakan menjadi dua kelas. Kelas pertama adalah batasan *exon to intron* dan kelas kedua adalah batasan *intron to exon.* Batasan DNA ini direpsesentasikan ke dalam sumbu *x* dan *y* pada koordinat Kartesius. Hal ini menyebabkan atribut untuk *dataset* 2 terdiri dari 2 buah yaitu *x* dan *y*. Sampel untuk *dataset* 1 berjumlah 594 *item* sedangkan *dataset* 2 berjumlah 854.

* 1. **Klasifikasi dengan metode *SVM***

Agar memenuhi syarat untuk diolah dengan menggunakan *SVM* *Gaussian,* data yang digunakan harus terlebih dahulu ditunjukkan tidak dapat dipisahkan secara linier. Suatu *dataset* yang tidak dapat dipisahkan secara linier akan terlihat pada grafik yang dihasilkan oleh dataset tersebut. Grafik untuk *dataset* 1 dan *dataset* 2 dapat dilihat pada Gambar 2:

**(a) (b)**

**Gambar 2.** Hasil Plot untuk *dataset* 1 (a) dan hasil plot untuk *dataset* 2 (b)

(Sumber: data olahan peneliti)

Gambar 2 bagian (a) dan (b) menunjukan bahwa baik *dataset* 1 maupun *dataset* 2 tidak dapat dipisahkan dengan menggunakan bidang pemisah linier. Karena data tidak dapat dipisahkan secara linier maka proses klasifikasi *SVM* akan menggunakan metode *kernel.*

Metode klasifikasi *SVM* memerlukanproses pelatihan dan pengujian. Menurut Keller (2013), perbandingan antara data yang digunakan untuk pelatihan dan pengujian tidak mengikat atau sesuai keinginan peneliti, namun pada umumnya digunakan perbandingan 70%-30%. Sebanyak 70% dari total data digunakan untuk proses pelatihan sedangkan sisanya sebanyak 30% digunakan dalam proses pengujian. Hal ini menyebabkan sebanyak 416 *item* dari *dataset* 1 dan 598 *item* dari *dataset* 2 digunakan untuk proses pelatihan metode *SVM* sedangkan sisanya sebanyak 178 *item* dari *dataset* 1 dan 256 *item* dari *dataset* 2 digunakan dalam proses pengujian keakuratan metode *SVM*. Tingkat keakuratan () diukur dengan rumus sebagai berikut:

dimana adalah total prediksi benar dan adalah total keseluruhan data yang diuji.

Proses klasifikasi *SVM* dilanjutkan dengan pelatihan untuk memperoleh nilai . *Software* Matlab versi 7 digunakan dalam proses pelatihan *SVM*. Setelah mendapatkan nilai , proses dilanjutakn dengan menghitung nilai *w* dan *b.* nilai *w* dan *b*  digunakan dalam membentuk fungsi pemisah. Nilai *w* dan *b* berbeda-beda untuk setiap *kernel* beserta parameter yang digunakan. Nilai *w* dan *b* berdasarkan jenis *kernel* dan parameternya untuk *dataset* 1 dan 2 dapat dilihat pada Tabel 1:

**Tabel 1.** Nilai *w* dan *b*berdasarkan jenis *kernel* dan parameternya untuk dataset 1 dan 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Jenis *Kernel* dan Parameternya | | Nilai *w* | | Nilai *b* | |
| *Dataset* 1 | *Dataset* 2 | *Dataset* 1 | *Dataset* 2 |
| *Gaussian* | = 0.5 | 7.8851  -14.5297 | -3.1465 1.5869 | 0.1112 | 1.9920 |
| = 0.1 | 50.9392  -93.2384 | -19.1234 7.0848 | 0.0593 | 0.2773 |
| = 1 | 5.4394  -7.2932 | -1.8653 1.3685 | 0.0690 | 1.7176 |
| = 10 | 58.8327  -97.3090 | -125.2087 13.4632 | 0.3600 | 0.0478 |
| Linier | | 2.7089  -3.3579 | -2.2068  0.1239 | -0.2457 | 0.2179 |
| Polinomial | Orde 2 | 2.5363  -3.2622 | -3.5365  -0.6228 | -0.0921 | -0.6991 |
| Orde 3 | 2.5359  -3.0608 | -3.6908  -0.2943 | 0.0285 | -0.2734 |

(Sumber: Data olahan peneliti)

Nilai w dan b pada Tabel 1 kemudian di subtitusi ke dalam persamaan (2.8) untuk memperoleh fungsi pemisah yang akan mengklasifikasi data ke dalam dua kelas.

* 1. **Grafik fungsi pemisah**

Setelah mendapatkan fungsi pemisah langkah selanjutnya adalah membentuk sebuah grafik fungsi pemisah. Grafik fungsi pemisah akan memisahkan data ke dalam dua kelas. Grafik yang dibentuk oleh tiap-tiap fungsi pemisah akan dijelaskan pada Tabel 2 berikut ini:

**Tabel 2.** Grafik fungsi pemisah yang dibentuk oleh setiap kernel beserta parameternya

|  |  |  |  |
| --- | --- | --- | --- |
| Jenis *Kernel* dan Parameternya | | Grafik fungsi Pemisah | |
| *Dataset* 1 | *Dataset* 2 |
| *Gaussian* | = 0.5 |  |  |
| = 0.1 |  |  |
| = 1 |  |  |
| = 10 |  |  |
| Linier | |  |  |
| Polinomial | Orde 2 |  |  |
| Orde 3 |  |  |

(Sumber: Data olahan peneliti)

* 1. **Kesalahan Klasifikasi**

Kesalahan klasifikasiadalah keadaan dimana suatu *item* data dikategorikan ke dalam kelas tertentu akan tetapi model klasifikasi *SVM* mengidentifikasi *item* data tersebut ke dalam kelas yang lain. Kesalahan klasifikasi terjadi dalam proses pengujian. Jumlah kesalahan klasifikasi untuk pengujian pada penelitian ini berbeda-beda sesuai dengan tipe *kernel* yang digunakan. Jumlah kesalahan klasifikasi dapat dilihat pada Tabel 3 dan Gambar 3 berikut ini:

**Tabel 3.** Jumlah kesalahan klasifikasi *dataset* 1 dan 2 untuk setiap *kernel* dan parameter yang digunakan.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Jenis *Kernel* dan Parameternya | | Jumlah kesalahan klaisfikasi | |
| *Dataset* 1 | *Dataset* 2 |
| 1 | *Gaussian* | = 0.5 | **9** | 19 |
| = 0.1 | 14 | **4** |
| = 1 | 10 | 65 |
| = 10 | 11 | 81 |
| 2 | Linier | | 116 | 80 |
| 3 | Polinomial | Orde 2 | 12 | 66 |
| Orde 3 | **9** | 65 |

(Sumber: Data olahan peneliti)

* + - * 1. **(b)**

**Gambar 3.** Jumlah kesalahan klasifikasi *dataset* 1 (a) dan *dataset* 2 (b) untuk setiap *kernel* beserta parameternya.

(Sumber: Data olahan peneliti)

Tabel 3 dan Gambar 3 menjelaskan bahwa *kernel* *Gaussian* dengan parameter = 0.5 dan kernel polinomial orde 3 memberikan tingkat kesalahan klasifikasi terkecil untuk *dataset* 1 (9) dan *kernel Gaussian* dengan parameter = 0.1 memberikan tingkat kesalahan yang paling kecil untuk *dataset* 2 (4). Pada *dataset* 1, *kernel* linier memberikan hasil kesalahan klasifikasi terbesar (116) sedangkan pada *dataset* 2 *kernel* *Gaussian* dengan parameter = 10 memberikan tingkat kesalahan tertinggi (81).

* 1. **Tingkat akurasi**

Tingkat akurasi diperolah dari jumlah data yang kelasnya diprediksi dengan tepat oleh model *SVM*. Tingkat akurasi pada penelitian ini berbeda-beda untuk setiap *kernel* yang digunakan. Tingkat akurasi setiap kernel dapat dilihat pada Tabel 4 dan Gambar 4:

**Tabel 4**. Tingkat akurasi setiap kernel yang digunakan pada *dataset* 1 dan *dataset* 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Jenis *Kernel* dan Parameternya | | Jumlah akurasi (%) | |
| *Dataset* 1 | *Dataset* 2 |
| 1 | *Gaussian* | = 0.5 | **94.94** | 92.58 |
| = 0.1 | 92.13 | **98.44** |
| = 1 | 94.38 | 74.61 |
| = 10 | 93.82 | 68.36 |
| 2 | Linier | | 34.83 | 68.75 |
| 3 | Polinomial | Orde 2 | 93.26 | 74.22 |
| Orde 3 | **94.94** | 75.78 |

(Sumber: Data olahan peneliti)

* + - * 1. (b)

**Gambar 4.** Grafik tingkat akurasi setiap *kernel* pada *dataset* 1 (a) dan *dataset* 2 (b) beserta parameternya.

(Sumber: Data olahan peneliti)

Berdasarkan Gambar 4, *kernel Gaussian* dengan parameter = 1 dan kernel polinomial orde 3 memiliki tingkat keakuratan yang paling tinggi (94.94%) untuk *dataset* 1. *Kernel Gaussian* dengan parameter = 0.1 mencapai tingkat keakuratan yang paling tinggi untuk *dataset* 2 (98.44%). *Kernel* linier untuk *dataset* 1 memiliki tingkat akurasi yang paling rendah (34.83%). Untuk *dataset* 2, tingkat akurasi terendah diperoleh dengan menggunakan *kernel Gaussian* dengan parameter =10 (68.36%).

1. **Kesimpulan dan Saran**
   1. **Kesimpulan**

Beberapa hal yang dapat disimpulkan dari penelitian ini adalah:

Klasifikasi dengan metode *SVM* dimulai dengan melakukan pelatihan untuk memperoleh fungsi pemisah yang membagi data ke dalam dua kelas. Setelah menemukan fungsi pemisah, proses selanjutnya adalah melakukan klasifikasi terhadap data uji untuk melihat tingkat akurasi model *SVM*.

*Kernel Gaussian* memiliki tingkat akurasi yang paling baik diantara *kernel lain* seperti polinomial atau linier untuk data tentang tinggi dan berat badan pemain basket dan *football* serta data tentang batas-batas dalam urutan DNA yang digunakan dalam penelitian ini.

Berdasarkan hasil pada Gambar 3 dan Gambar 4, ada dua hal utama yang mempengaruhi besarnya tingkat kesalahan klasifikasi dan tingkat akurasi klasifikasi dengan metode SVM. Kedua hal tersebut adalah pemilihan *kernel* beserta parameternya. Pemilihan parameter dan parameter yang tepat akan memberikan hasil kesalahan klasifikasi yang lebih sedikit.

* 1. **Saran**

Beberapa hal yang diusulkan untuk kelanjutan penelitian ini adalah sebagai berikut:

1. Mengembangkan metode *multiclass* *SVM* sehingga suatu data dapat diklasifikasikan ke dalam lebih dari 2 kelas.
2. Menemukan suatu metode untuk menentukan parameter terbaik untuk *kernel* *Gaussian* dimana dengan parameter tersebut diharapkan tingkat akurasi dari *kernel* *Gaussian* menjadi maksimal.
3. Mengembangkan metode *SVM* yang dipadukan dengan lebih dari satu *kernel* (*multi* *kernel*) dengan tujuan meningkatkan tingkat akurasi klasifikasi data dengan metode *SVM*.

**DAFTAR PUSTAKA**

1. Abe, Shigeo. 2006. *Support Vector Machines for Pattern Classification.* United Kingdom: Springler.
2. Albatal, dan Little. 2013. Empirical exploration of extreme *SVM*-RBF parameter values for visual object classification*.* *Lecture Notes in Computer Science,* **8326**(14):299-306.
3. Borivikov, E, A. 1999. *An Evaluation of Support Vector Machines as a Pattern Recognition Tool.* Maryland: University of Maryland.
4. Boswell, D. 2002. *Introduction to Support Vector Machines.* California: California Institute of Technology.
5. Burgess, J. 1998. A Tutorial on Support Vector Machines for Pattern Recognition*. Data Mining and Knowledge Discovery,* **2**(2):121-167.
6. Charu, Philip. 1999. Methodologies for Knowledge Discovery and Data Mining. *Lecture Notes in Computer Science,* **1574**(1999):13-23.
7. Christianini, N, dan Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge : University Press.
8. DNA *Dataset*.

URL: [http://www.csie.ntu.edu.tw/~cjlin/lib*SVM*tools/*dataset*s/binary.html #fourclass](http://www.csie.ntu.edu.tw/~cjlin/libSVMtools/datasets/binary.html%20#fourclass) diakses pada 28 Mei 2014

1. Fadli, A. 2008. *Aplikasi SVM dalam teknik ekstrasi informasi.* Jakarta : Universitas Gunadarma.
2. Gunn, S. 1998. *Support Vector Machines for Classification and Regression*. Southampton: University of Southampton
3. Hsu, Chih-Wei *et al*. 2010. *A Practical Guide to Support Vector Classification.* Taipei: National Taiwan University.
4. Hofmann, M. 2006. *Support Vector Machines — Kernels and the Kernel Trick.* Bamberg: University Of Bamberg.
5. Keller, F. 2013. *Connectionist and Statistical Language Processing*. Saarbrücken : Saarland University.
6. Nugroho, A.M. 2008. *Pengantar Support Vector machine.* Pusat Teknologi Informasi dan komunikasi BPP Teknologi.
7. Chicago Sports. 2014. *Football Player Database.*

URL:<http://chicagosports.sportsdirectinc.com/football/nfl-teams.aspx?page>= /data/nfl/teams/rosters/roster2.html (Diakses pada tanggal 11 Mei 2014 Pukul 16.00).

1. Snijders, C. 2012. “Big data”: Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science,* **7**(1):1-5.
2. Santoso, H. 2012. *Analisis dan Prediksi pada Perilaku Mahasiswa Diploma untuk Melanjutkan Studi Ke Jenjang Sarjana Menggunakan Teknik Decision Tree dan Support Vektor Machine*. Medan: Universitas Sumatera Utara.
3. Sportilustrated. 2014. *NBA Player Statistic.*

URL: <http://www.sportsillustrated.cnn.com/basketball/nba/teams/>

(Diakses pada 11 Mei 2014 pukul 16.09).