

**KLASIFIKASI WILAYAH DESA-PERDESAAN DAN
DESA-PERKOTAAN WILAYAH KABUPATEN SEMARANG
DENGAN *SUPPORT VECTOR MACHINE* (SVM)**



=====
SKRIPSI
=====

**Disusun Oleh :
MEKAR SEKAR SARI
NIM. 24010210120008**

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2014**

**KLASIFIKASI WILAYAH DESA-PERDESAAN DAN
DESA-PERKOTAAN WILAYAH KABUPATEN SEMARANG
DENGAN *SUPPORT VECTOR MACHINE* (SVM)**

**MEKAR SEKAR SARI
NIM. 24010210120008**

Skripsi

Diajukan Sebagai Syarat untuk Mendapatkan Gelar Sarjana
Pada Jurusan Statistika

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2014**

LEMBAR PENGESAHAN I

Judul : Klasifikasi Wilayah Desa-Perdesaan Dan Desa-Perkotaan
Wilayah Kabupaten Semarang Dengan *Support Vector Machine*
(SVM)

Nama : Mekar Sekar Sari

NIM : 24010210120008

Telah diujikan pada sidang Tugas Akhir tanggal 19 September 2014 dan dinyatakan lulus pada tanggal 29 September 2014.

Semarang, September 2014

Mengetahui,

Ketua Jurusan Statistika

Fakultas Sains dan Matematika Undip

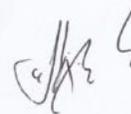


Dra. Dwi Isprivanti, M.Si

NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir

Ketua,



Dra. Suparti, M.Si.

NIP. 196509131990032001

LEMBAR PENGESAHAN II

Judul : Klasifikasi Wilayah Desa-Perdesaan Dan Desa-Perkotaan
Wilayah Kabupaten Semarang Dengan *Support Vector Machine*
(SVM)

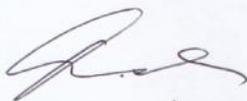
Nama : Mekar Sekar Sari

NIM : 24010210120008

Telah diujikan pada sidang Tugas Akhir tanggal 19 September 2014.

Semarang, September 2014

Pembimbing I



Diah Safitri, S.Si,M.Si.

NIP. 197510082003122001

Pembimbing II



Sugito, S.Si,M.Si.

NIP. 197610192005011001

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat, berkah, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Klasifikasi Wilayah Desa-Perdesaan Dan Desa-Perkotaan Wilayah Kabupaten Semarang Dengan *Support Vector Machine* (SVM)”**.

Tanpa bantuan dan dukungan dari berbagai pihak, penulis tidak akan mampu menyelesaikan laporan ini. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Ibu Dra. Hj. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Ibu Diah Safitri, S.Si, M.Si selaku dosen pembimbing I dan Bapak Sugito, S.Si, M.Si selaku dosen pembimbing II.
3. Bapak Ibu Dosen Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
4. Semua pihak yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari masih terdapat kekurangan pada Laporan Tugas Akhir ini. Oleh karena itu, saran dan kritik dari semua pihak yang bersifat membangun sangat penulis harapkan demi kesempurnaan Tugas Akhir ini.

Semarang, September 2014

Penulis

ABSTRAK

Pada penelitian ini akan dilakukan klasifikasi wilayah berdasarkan status desa-perdesaan dan desa-perkotaan yang mengacu pada perbedaan karakteristik/kondisi antarwilayah di Indonesia dengan metode *Support Vector Machine* (SVM). Klasifikasi pada permasalahan ini bekerja dengan membangun fungsi pemisah yang melibatkan fungsi kernel untuk memetakan data input ke ruang dimensi yang lebih tinggi. Algoritma *Sequential Minimal Optimization* (SMO) digunakan pada proses *training* data klasifikasi wilayah desa-perdesaan dan desa-perkotaan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal. Untuk menentukan fungsi kernel dan parameter yang sesuai dengan data pada penelitian ini digunakan metode *grid search* yang dipadukan dengan metode *leave-one-out cross-validation*. Pada klasifikasi menggunakan metode SVM ini didapatkan ukuran ketepatan klasifikasi, yaitu nilai akurasi terbaik sebesar 90% dengan menggunakan fungsi kernel *Radial Basis Function* (RBF) dengan parameter $C=100$ dan $\gamma=2^{-5}$.

Kata Kunci : Klasifikasi, *support vector machine*, *sequential minimal optimization*, *grid search*, *leave-one-out*, *cross-validation*, desa-perdesaan, desa-perkotaan

ABSTRACT

This research will be carry out classification based on the status of the rural and urban regions that reflect the differences in characteristics/ conditions between regions in Indonesia with Support Vector Machine (SVM) method. Classification on this issue is working by build separation functions involving the kernel function to map the input data into a higher dimensional space. Sequential Minimal Optimization (SMO) algorithms is used in the training process of data classification of rural and urban regions to get the optimal separation function (hyperplane). To determine the kernel function and parameters according to the data, grid search method combined with the leave-one-out cross-validation method is used. In the classification using SVM, accuracy is obtained, which the best value is 90% using Radial Basis Function (RBF) kernel functions with parameters $C=100$ dan $\gamma=2^{-5}$.

Keywords : classification, support vector machine, sequential minimal optimization, grid search, leave-one-out, cross validation, rural, urban

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN I.....	ii
LEMBAR PENGESAHAN II.....	iii
KATA PENGANTAR	iv
ABSTRAK	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR SIMBOL.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.3 Batasan Masalah.....	3
1.4 Tujuan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Klasifikasi Wilayah Desa-Perdesaan Dan Desa-Perkotaan	5
2.2 <i>Support Vector Machine (SVM)</i>	10
2.2.1 Konsep SVM.....	10
2.2.2 SVM pada <i>Linearly Separable Data</i>	13
2.2.3 SVM pada <i>Non-linearly Separable Data</i>	16
2.2.4 Metode Kernel dan Klasifikasi Nonlinier pada SVM	17

2.3	Optimasi <i>Hyperplane</i> SVM	19
2.4	Seleksi Parameter	23
2.4.1	<i>Cross-Validation</i>	23
2.4.2	<i>Grid Search</i>	24
2.5	Pengukuran Kinerja Klasifikasi	24
BAB III METODOLOGI PENELITIAN.....		26
3.1	Jenis dan Sumber Data	26
3.2	Variabel Penelitian	26
3.3	Langkah-langkah Analisis.....	28
3.4	Diagram Alir	30
BAB IV PEMBAHASAN		31
4.1	Analisis Deskriptif	31
4.2	Klasifikasi dengan SVM	32
BAB V KESIMPULAN		48
DAFTAR PUSTAKA		49
LAMPIRAN.....		51

DAFTAR GAMBAR

	Halaman
Gambar 2.1 <i>Hyperplane</i> terbaik.....	11
Gambar 2.2 <i>Hyperplane</i> terbaik dengan margin maksimal	12
Gambar 3.1 Diagram Alir Klasifikasi dengan Metode SVM.....	30
Gambar 4.1 Persentase Jumlah Desa-Perdesaan dan Desa-Perkotaan Kabupaten Semarang	31

DAFTAR TABEL

	Halaman
Tabel 2.1 Variabel, Klasifikasi, Skor, dan Kriteria Desa Perkotaan	7
Tabel 2.2 Matriks Konfusi.....	25
Tabel 4.1 Jumlah Wilayah Administratif Terkecil Kabupaten Semarang Berdasarkan Status Daerah	31
Tabel 4.2 Persentase Akurasi Hasil Klasifikasi Dengan <i>Leave-One-Out Cross-Validation</i>	34
Tabel 4.3 Error Klasifikasi Untuk Setiap Parameter <i>cost</i> (C)	35
Tabel 4.4 Error Klasifikasi Untuk Setiap Parameter <i>degree</i> (d) dan <i>cost</i> (C).....	37
Tabel 4.5 Nilai $\alpha_i y_i$ untuk <i>hyperplane</i> dengan fungsi kernel <i>Polynomial</i> dengan parameter C=0,01 dan d=2	38
Tabel 4.6 <i>Support vector</i> untuk <i>hyperplane</i> dengan fungsi kernel <i>Polynomial</i> dengan parameter C=0,01 dan d=2.....	39
Tabel 4.5 Error Klasifikasi Untuk Setiap Parameter <i>gamma</i> (γ) dan <i>cost</i> (C).....	41
Tabel 4.7 Nilai $\alpha_i y_i$ untuk <i>hyperplane</i> dengan fungsi kernel RBF dengan parameter C=100 dan $\gamma = 2^{-5}$	42
Tabel 4.8 <i>Support vector</i> untuk <i>hyperplane</i> dengan fungsi kernel RBF dengan parameter C=100 dan $\gamma = 2^{-5}$	43
Tabel 4.6 Akurasi Hasil Klasifikasi Dengan Metode SVM	47

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Klasifikasi Wilayah Desa-perdesaan Desa-Perkotaan Kabupaten Semarang.....	51
Lampiran 2 Pembagian Data <i>Training-Testing</i> Secara Acak.....	61
Lampiran 3 Syntax R Klasifikasi dengan SVM.....	65
Lampiran 4 Output Klasifikasi SVM Dengan Kernel Linier	66
Lampiran 5 Output Klasifikasi SVM Dengan Fungsi Kernel <i>Polynomial</i>	68
Lampiran 6 Output Klasifikasi SVM Dengan Fungsi Kernel <i>Radial Basis Function (RBF)</i>	72

DAFTAR SIMBOL

x_i	: data input ke-i
$f(x)$: fungsi <i>hyperplane</i>
$g(x)$: label kelas data input x
w	: parameter fungsi <i>hyperplane</i>
b	: parameter fungsi <i>hyperplane</i>
R^n	: ruang vector dengan dimensi n
y_i	: label kelas data ke-i
z	: data uji
$\ w\ $: norm vektor w
d	: jarak antara dua <i>hyperplane</i>
α_i	: alpha, pengali <i>Lagrange</i> yang berkorespondensi dengan x_i
L_p	: fungsi <i>Lagrange</i> primal
L_d	: fungsi <i>Lagrange</i> dual
C	: parameter <i>cost</i>
ξ_i	: variable <i>slack</i> ke-i
$K(x_i, x_j)$: fungsi Kernel
$\Phi(x_i)$: transformasi data input (x_i) di <i>future space</i>
c	: konstanta fungsi Kernel <i>Polynomial</i>
γ	: gamma, parameter fungsi Kernel <i>Radial Basis Function</i> (RBF)
σ	: sigma, parameter fungsi Kernel <i>Tangent hyperbolic</i> (sigmoid)
H	: Batas atas nilai pengali <i>Lagrange</i>

L : Batas bawah nilai pengali *Lagrange*

s : *dot-product* dua elemen vektor

BAB I

PENDAHULUAN

1.1 Latar Belakang

Wilayah Indonesia dibagi ke dalam beberapa tingkat wilayah administratif, yaitu provinsi, kabupaten/kota, kecamatan, dan desa atau disebut dengan nama lain yang merupakan wilayah administratif terkecil (Badan Pusat Statistik, 2010). Untuk berbagai keperluan, data mengenai klasifikasi wilayah desa dan kota sangat bermanfaat terutama dalam hal perencanaan pembangunan. Badan Pusat Statistik (BPS) merupakan salah satu badan penyedia data pembagian wilayah administratif yang telah diklasifikasikan menjadi wilayah desa-perdesaan dan desa-perkotaan untuk keperluan statistik dan keperluan lainnya yang berhubungan dengan analisis dan perencanaan pembangunan. Sebagai contoh, BPS biasanya menggunakan klasifikasi desa perkotaan perdesaan sebagai dasar untuk merencanakan kegiatan sensus atau survey (Badan Pusat Statistik, 2010).

Perencanaan pembangunan wilayah mencakup berbagai aspek yang tentunya mempertimbangkan peran keterkaitan antara desa dan kota. Mike Douglass (1998) dalam Tarigan (2003), melalui konsep agropolitan menekankan bahwa pengembangan desa dapat tercapai dengan baik apabila desa tersebut dikaitkan dengan pengembangan kota dalam wilayah tersebut. Keterkaitan antara desa dan kota juga harus dipandang dari hal budaya dan interaksi sosial yang memungkinkan adanya timbal balik yang saling menguntungkan.

Klasifikasi antara wilayah desa dan kota sangat penting dilakukan untuk menentukan jenis intervensi apa yang akan diberikan. Kedua wilayah tersebut memiliki interdependensi yang tinggi dalam rantai keterkaitan permintaan dan penawaran. Di samping pertimbangan ekonomi seperti sudah diuraikan sebelumnya, keterkaitan antara kedua wilayah tersebut juga penting untuk mengatasi masalah urbanisasi yang memiliki implikasi politik. Karenanya, keterkaitan desa kota tidak sekedar membawa implikasi ekonomi tetapi juga dampak politik (Tarigan, 2003).

Penggolongan wilayah desa-perdesaan dan desa-perkotaan oleh BPS melibatkan beberapa variabel yang telah ditetapkan berdasarkan Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010 tentang Klasifikasi Perkotaan dan Perdesaan di Indonesia. Mengacu pada kriteria desa-perdesaan dan desa-perkotaan yang tercantum dalam peraturan tersebut penulis bermaksud melakukan klasifikasi dengan menerapkan ilmu statistika menggunakan teknik *data mining*.

Menurut Santosa (2007), *data mining*, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Terdapat beberapa teknik *data mining* yang dapat digunakan dalam permasalahan klasifikasi. Namun, pada permasalahan ini penulis akan menggunakan teknik klasifikasi *Support Vector Machine* (SVM) untuk mengetahui ketepatan hasil klasifikasi melalui nilai akurasi yang diperoleh.

Support Vector Machine (SVM) adalah suatu teknik untuk menemukan fungsi pemisah (*hyperplane*) yang bisa memisahkan dua himpunan data dari dua

kelas yang berbeda (Vapnik, 1995). Ide dasar SVM adalah memaksimalkan batas *hyperplane*. *Hyperplane* dengan margin yang maksimal akan memberikan generalisasi yang lebih baik pada metode klasifikasi. SVM bekerja dengan baik pada himpunan data berdimensi tinggi (Prasetyo, 2012). Mengaplikasikan metode klasifikasi SVM pada permasalahan tersebut maka penulis mengambil “Klasifikasi Wilayah Desa-Perdesaan dan Desa-Perkotaan Wilayah Kabupaten Semarang Dengan *Support Vector Machine*” sebagai judul Tugas Akhir.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang, rumusan masalah yang akan dibahas dalam penelitian ini adalah mengaplikasikan metode klasifikasi *Support Vector Machine* (SVM) pada data desa-perdesaan dan desa-perkotaan di Kabupaten Semarang untuk mengetahui ketepatan hasil klasifikasi metode tersebut melalui nilai akurasi yang diperoleh.

1.3 Batasan Masalah

Batasan ruang lingkup permasalahan dari penelitian ini adalah sebagai berikut :

1. Data yang digunakan merupakan data klasifikasi desa-perdesaan dan desa-perkotaan wilayah Kabupaten Semarang.
2. Dalam penelitian ini terdiri dari sepuluh variabel independen yang merupakan kriteria klasifikasi desa-perdesaan dan desa-perkotaan menurut Badan Pusat Statistik.

3. Fungsi kernel yang digunakan untuk menentukan *hyperplane* terbaik pada klasifikasi menggunakan metode SVM adalah fungsi kernel linier, *polynomial* dan *Radial Basis Function* (RBF).

1.4 Tujuan

Tujuan dari penelitian ini, antara lain :

1. Menentukan fungsi kernel dan parameter terbaik untuk *hyperplane* SVM dengan menerapkan metode *grid search*.
2. Menentukan *hyperplane* dengan parameter terbaik yang telah didapatkan untuk klasifikasi data menggunakan SVM.
3. Evaluasi hasil klasifikasi dengan menghitung akurasi hasil prediksi klasifikasi dengan metode SVM.