

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

NAMA : DINDA FIRLY PARAMITHA
NRP : 5110100084
DOSEN WALI : Ahmad Saikhu, S.Si., M.T.
DOSEN PEMBIMBING : 1. Arya Yudhi Wijaya, S.Kom, M.Kom.
2. Rully Soelaiman, S.Kom, M.Kom.

2. JUDUL TUGAS AKHIR

“Implementasi *Support Vector Machine* dengan Optimasi *Pruning* pada Persoalan Klasifikasi Pola”

3. LATAR BELAKANG

Support Vector Machine (SVM) merupakan topik penting dalam banyak area seperti pengenalan pola, pengolahan citra, *machine learning* dan *bioinformatics*. Hal ini dikarenakan SVM dapat diaplikasikan secara luas. SVM merupakan sebuah pendekatan untuk menemukan *hyperplane* yang dapat memaksimalkan margin dan meminimalkan kesalahan klasifikasi ketika diberikan masalah dua kelas secara linear dipisahkan.

Pendekatan SVM dapat dibagi menjadi dua yaitu dari pandangan aljabar dan pandangan geometri. Pandangan aljabar memiliki tujuan bagaimana meminimalkan kesalahan klasifikasi dan mengurangi biaya komputasi dengan *Sequential Minimal Optimization* (SMO) [1], SVM dengan *Soft Margin* [2], v-SVM [3], Kernel SVM [4], dan *Support Vector Regression Machine* [5]. Lain halnya dengan pandangan geometri yang berguna untuk menyelesaikan masalah klasifikasi.

SVM memiliki beberapa kelemahan. Salah satunya adalah sulitnya pemakaian SVM apabila menyelesaikan masalah yang berskala besar. Maksud dari pengertian skala besar dalam hal ini adalah berkaitan dengan jumlah *sample* yang diolah. Selain

itu, kelemahan SVM yang lainnya adalah hanya bisa menyelesaikan masalah klasifikasi dengan dua kelas.

Untuk menyelesaikan masalah SVM agar dapat digunakan sebagai pengklasifikasi banyak kelas, maka dalam Tugas Akhir ini akan diusulkan sebuah metode yang dapat menyelesaikan permasalahan tersebut berdasarkan properti dari *support vectors* yang dapat mengeliminasi *training vectors* duplikat pada waktu bersamaan. Metode yang dipilih adalah SVM dengan menggunakan SMO yang dikembangkan dengan proses *pruning*. Hasil yang diharapkan adalah keakuratan dari metode yang dikembangkan apabila diaplikasikan terhadap segmentasi citra maupun klasifikasi, serta dapat mengurangi biaya komputasi.

4. RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana memahami metodologi *Support Vector Machine* dengan proses *pruning*?
2. Bagaimana melakukan praproses data agar data dapat diaplikasikan dengan baik?
3. Bagaimana melakukan eliminasi *training vectors* dengan menggunakan metode *pruning*?
4. Bagaimana menggunakan metode *Support Vector Machine* dengan proses *pruning*?

5. BATASAN MASALAH

Adapun batasan ruang lingkup permasalahan dari pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Implementasi dilakukan dengan menggunakan Matlab 7.6.
2. Modifikasi *Support Vector Machine* dengan menggunakan metode *pruning*.
3. Data set yang digunakan adalah data citra untuk diaplikasikan pada segmentasi citra dan data dari UCI untuk diaplikasikan pada klasifikasi.

6. TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan dari pembuatan Tugas Akhir ini adalah sebagai berikut :

1. Mengimplementasikan metode *Support Vector Machine* dengan proses *pruning*.
2. Melakukan praproses data agar data dapat diaplikasikan dengan baik.
3. Melakukan eliminasi *training vectors* dengan menggunakan metode *pruning*.
4. Menggunakan metode *Support Vector Machine* dengan proses *pruning*.

7. MANFAAT TUGAS AKHIR

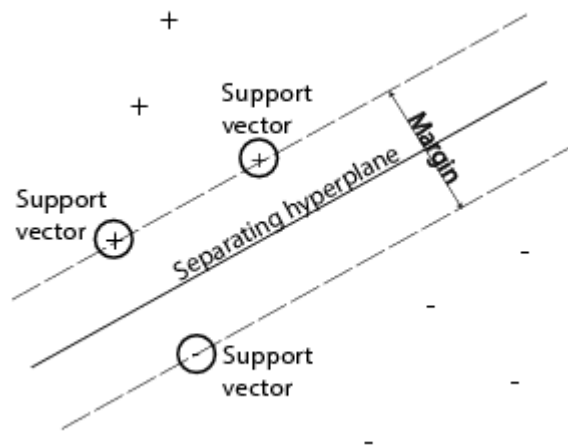
Tugas Akhir ini dikerjakan dengan harapan dapat memberikan manfaat dalam bidang informatika dan dapat memberikan sumbangan pemikiran dalam mempercepat performa *Support Vector Machine* (SVM) yang dimodifikasi dengan proses *pruning* untuk diaplikasikan pada klasifikasi dan segmentasi citra.

8. TINJAUAN PUSTAKA

8.1 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) merupakan salah satu sistem cerdas yang dibangun berdasarkan teori statistik. SVM merupakan metode untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Tujuan utama SVM adalah untuk mendapatkan keputusan yang memiliki jarak margin paling jauh dari titik data pelatihan yang terdekat. Prinsip dasar klasifikasi SVM yang pertama adalah memetakan masukan yang berupa vektor ke dalam ruang fitur (*feature space*) secara linear maupun non-linear dan sesuai dengan jenis kernel yang dipakai.

Support vectors merupakan titik data yang paling dekat dengan *hyperplane* yang memisahkan. Titik-titik ini berada pada perbatasan lempengan. Hal ini dijelaskan dengan (+) menunjukkan titik data dari tipe 1, dan (-) menunjukkan titik data tipe -1 seperti yang diilustrasikan pada Gambar 1 [6].



Gambar 1 Ilustrasi *Support Vectors* yang Berada di Perbatasan

Performa klasifikasi SVM bergantung pada kombinasi beberapa parameter antara lain kapasitas parameter C, tipe kernel K, dan beberapa parameter lain yang berkoresponden. Pada teori optimasi SVM biasanya melibatkan beberapa parameter. Salah satunya adalah parameter C yang mengontrol hasil keluaran agar margin maksimal dan kesalahan minimal. Parameter lain terbentuk pada pemetaan non-linear ruang fitur yang disebut parameter kernel. Secara sederhana proses yang dilakukan biasanya memungkinkan parameter C dikenal sebagai parameter kernel sehinggalah semua parameter dapat diperlakukan sebagai satu kerangka [7].

8.2 *Sequential Minimal Optimization* (SMO)

Sequential Optimization Minimal (SMO) ditemukan oleh John Platt pada tahun 1998. SMO merupakan sebuah algoritma sederhana yang cepat dalam menyelesaikan masalah SVM *Quadratic Problem* (QP) tanpa memakai penyimpanan matriks ekstra dan tanpa melibatkan sebuah kegiatan numerikal berulang pada setiap sub masalah. SMO menguraikan keseluruhan masalah QP menjadi sub masalah QP yang mirip dengan metode Otsuna.

Berbeda dengan metode sebelumnya, SMO memilih untuk menyelesaikan kemungkinan masalah optimasi terkecil pada setiap langkahnya. Untuk standar masalah SVM QP, kemungkinan masalah optimasi terkecil melibatkan dua Lagrange *multipliers* karena Lagrange *multipliers* harus memenuhi pembatas kesetaraan linear. Pada setiap langkahnya, SMO memilih dua Lagrange *multipliers* untuk dioptimasi bersama. Setelah itu, SMO menemukan nilai-nilai optimal untuk *multipliers* ini dan memperbarui SVM untuk menggambarkan nilai-nilai optimal baru.

SMO memiliki 3 komponen yaitu metode analitis untuk menyelesaikan masalah Lagrange *multipliers*, sebuah heuristik untuk memilih *multipliers* manakah untuk mengoptimalkan, dan metode untuk menghitung. Keuntungan dari SMO adalah dalam menyelesaikan masalah dua Lagrange *multipliers* dapat dilakukan secara analitis. Hal tersebut menyebabkan seluruh iterasi dalam karena optimasi numerik QP dapat dihindari [8].

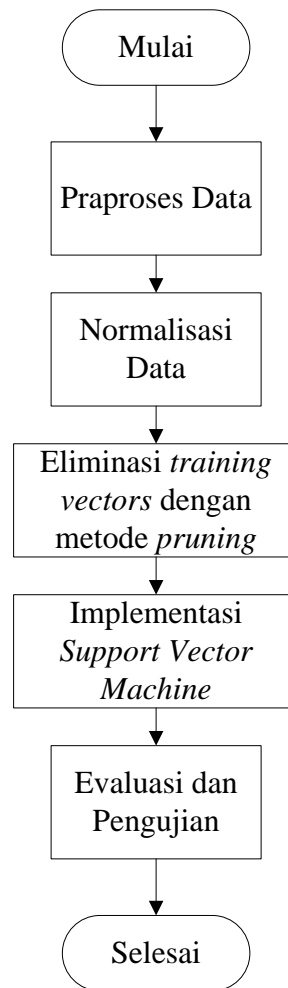
8.3 Proses *Pruning*

Proses *pruning* merupakan metode *machine learning* yang berguna untuk mengurangi ukuran dari keputusan dengan menghapus bagian yang kurang memberikan informasi yang dibutuhkan dalam mengklasifikasikan kasus. Tujuan dari proses *pruning* adalah mengurangi kompleksitas dari *classifier* akhir. Selain itu, proses *pruning* juga membuat akurasi prediksi menjadi lebih baik dengan melakukan pengurangan *overfitting* dan penghapusan bagian *classifier* yang mungkin didasarkan dari data yang *noise* atau salah.

Pemangkasan dilakukan tanpa mengurangi akurasi prediksi yang diukur dengan separangkat tes atau menggunakan *cross-validation*. Untuk mengoptimalkan kinerja, ada banyak teknik yang dapat digunakan untuk melakukan pemangkasan [9].

9. RINGKASAN ISI TUGAS AKHIR

Topik Tugas Akhir yang akan dikerjakan berkaitan dengan pengaplikasian *Support Vector Machine* (SVM) dengan menggunakan proses *pruning* terhadap segmentasi citra dan klasifikasi. Proses-proses yang dilakukan dalam Tugas Akhir ini akan dijelaskan dalam diagram alur Gambar 2.



Gambar 2 Diagram Alur untuk Pengerjaan Tugas Akhir

Proses pertama yang dilakukan adalah praproses data yaitu mengolah data agar dapat diimplementasikan dengan baik. Setelah itu melakukan normalisasi data yang digunakan agar data yang diproses tidak memiliki jarak yang terlalu jauh antara satu dengan lainnya. Proses berikutnya adalah melakukan eliminasi *training vectors* dengan proses *pruning*. Hal ini diperlukan untuk mencegah adanya data yang sama sehingga menyebabkan lamanya komputasi.

Setelah proses *pruning*, barulah menjalankan SVM dengan SMO sebagai pengklasifikasi *training*. Kemudian dilakukanlah pengevaluasian dan pengujian untuk mengetahui tingkat akurasi metode tersebut apabila diaplikasikan terhadap data citra untuk segmentasi maupun data UCI untuk klasifikasi.

10.METODOLOGI

a. Penyusunan proposal tugas akhir

Penyusunan proposal tugas akhir merupakan tahap awal dalam proses pengerjaan Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan implementasi *support vector machine* dengan optimasi *pruning* pada persoalan klasifikasi pola.

b. Studi literatur

Pada tahap ini dilakukan pencarian informasi dan studi literatur yang dibutuhkan terhadap SVM, SMO, metode *pruning*. Informasi akan diperoleh dari buku acuan maupun *internet*.

c. Implementasi

Implementasi merupakan tahap untuk membangun algoritma tersebut. Untuk membangun algoritma yang telah dirancang sebelumnya, diimplementasikan dengan menggunakan MATLAB.

d. Pengujian dan evaluasi

Pada tahap ini dilakukan uji coba dengan menggunakan data citra dan data dari UCI untuk mencoba jalannya aplikasi apakah telah sesuai dengan rancangan desain implementasi yang dibuat, serta untuk mencari kesalahan-kesalahan program yang mungkin terjadi untuk selanjutnya dilakukan penyempurnaan. Metode evaluasi yang akan digunakan dalam Tugas Akhir ini adalah dengan menggunakan metode *Support Vector Machine* yang dikembangkan dengan proses *pruning*.

e. Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
 - a. Latar Belakang
 - b. Rumusan Masalah
 - c. Batasan Tugas Akhir
 - d. Tujuan
 - e. Metodologi
 - f. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

11. JADWAL KEGIATAN

Berikut ini merupakan jadwal kegiatan Tugas Akhir yang akan dijelaskan dalam Tabel 1.

Tabel 1 Jadwal Kegiatan Tugas Akhir

Tahapan	2014																			
	Februari				Maret				April				Mei				Juni			
Penyusunan Proposal																				
Studi Literatur																				
Implementasi																				
Pengujian dan evaluasi																				
Penyusunan buku																				

12. DAFTAR PUSTAKA

- [1] John C. Platt, "Using analytic QP and sparseness to speed training of support vector," *NIPS*, pp. 557-563, 1998.
- [2] Corina Cortes and Vladimir Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, September 1995.
- [3] C.C. Chang and C.J. Lin, "Training v-support vector classifiers: theory and algorithms," *Neural Computation*, vol. 13, no. 9, pp. 2119-2147, 2001.
- [4] Bernhard Schölkopf, Patrice Simard, Alex J. Smola, and Vladimir Vapnik, "Prior knowledge in support vector kernels," *NIPS*, pp. 640-646, 1997.
- [5] Harris Drucker, Christopher J.C. Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik, "Support vector regression machines," *NIPS*, pp. 155-161, 1996.
- [6] Mathworks. (2014, Februari) Support Vector Machines (SVM) - MATLAB & Simulink. [Online]. <http://www.mathworks.com/help/stats/support-vector-machines-svm.html>
- [7] Wikipedia. (2014, Februari) Support vector machine - Wikipedia, the free encyclopedia. [Online]. http://en.wikipedia.org/wiki/Support_vector_machine
- [8] John C. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," *Microsoft Research*, 1998.
- [9] Wikipedia. (2014, Februari) Pruning (decision trees) - Wikipedia, the free encyclopedia. [Online]. [http://en.wikipedia.org/wiki/Pruning_\(decision_trees\)](http://en.wikipedia.org/wiki/Pruning_(decision_trees))