APLIKASI *SUPPORT VECTOR MACHINES* (SVM)

PADA DATA BIOINFORMATIKA

(Studi kasus : Klasifikasi pada Data *Maize Growth Stages*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

Jurusan Statistika

Leni Anggraini Susanti

14611178

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

2018

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Aplikasi *Support Vector Machines (SVM)* pada Data Bionformatika
Nama Mahasiswa	:	Leni Anggraini Susanti
NIM	:	14 611 178

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, 31 Juli 2018

Menyetujui,	**Menyetujui,**
Dosen Pembimbing I	**Dosen Pembimbing II**
Dr. Techn. Rohmatul Fajriyah, S.Si., M.Si	**Husna Nugrahapraja, Ph.D**

HALAMAN PENGESAHAN

TUGAS AKHIR

APLIKASI *SUPPORT VECTOR MACHINES* (SVM)
PADA DATA BIOINFORMATIKA

Nama Mahasiswa : Leni Anggraini Susanti

Nomor Mahasiswa : 14 611 178

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 31 JULI 2018

Nama Penguji	Tanda Tangan
1. Muryanto, S.P., M.Si	.....
2. Muh. Hasan Sidiq Kurniawan, M.Sc	.....
3. Husna Nugrahapraja, Ph.D	.....
4. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si	.....

Mengetahui,
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

(Prof. Riyanto, S.Pd., M.Si., Ph.D)

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Assalamu'alaikum Wr.Wb.

Alhamdulillahirabbil'alamin. Puji Syukur kehadirat Allah SWT, yang telah melimpahkan rahmat, taufik serta hidayah-Nya atas segala yang telah diberikan kepada penulis. Shalawat serta salam juga tak henti-hentinya kita curahkan kepada junjungan Nabi Besar Muhammad SAW beserta seluruh keluarga beliau, sahabat beliau, para pejuang dan para ulama hingga akhir zaman.

Penyusunan Tugas Akhir yang berjudul "**Aplikasi *Support Vector Machines* (SVM) Pada Data Bioinformatika**" dengan studi kasus data *maize growth stages* ini merupakan salah satu persyaratan yang harus dipenuhi dalam memperoleh gelar sarjana Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Dalam penyusunan skripsi ini penulis banyak mmperoleh bantuan dari berbagai pihak, baik berupa bimbingan, ilmu, saran hingga kritik yang membangun. Oleh karena itu pada kesempatan ini penulis ingin mengungkapkan rasa terima kasih kepada :

1. Orang tua dan keluarga tercinta yang senantiasa selalu memberikan dukungan baik secara moril dan materiil kepada penulis untuk selalu berdoa dalam lindungan-Nya dan selalu berusaha dalam mencapai yang terbaik.
2. Bapak Fathul Wahid, S.T., M.Sc., Phd, selaku Rektor Universitas Islam Indonesia.
3. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
4. Bapak RB Fajriya Hakim, M.Si., selaku Ketua Jurusan Statistika beserta seluruh jajarannya.
5. Ibu Dr.techn.Rohmatul Fajriyah, S.Si., M.Si, dan Bapak Husna Nugrahapraja, Ph.D selaku dosen pembimbing Tugas Akhir atas

6. bimbingan, waktu, tenaga, ilmu, arahan dan masukan kepada penulis sehingga dapat menyelesaikan Tugas Akhir dengan baik.
7. Sahabat-sahabat penulis yang juga seperjuangan Bila, Ridha, Sita, Panji, Aulia baik yang telah membantu, menemani maupun memberi dukungan kepada penulis selama pengerjaan tugas akhir ini.
8. Tim SBRC yaitu teman-teman sepembimbing dan seperjuangan ummi, lina, ika, hawila, melda, maidah, nanda, ucen, gilang, ari, shodiq dan aat yang selalu ada dan selalu mendukung serta memberi motivasi.
9. Keluarga Besar Statistika UII secara umum dan keluarga besar Statistika angkatan 2014 secara khusus. Terimakasih semuanya.
10. Keluarga Unit Purworejo 91 Kuliah Kerja Nyata, terima kasih atas pelajaran dan dukungan yang sangat berharga selama KKN dan sampai saat ini.
11. Semua pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Semoga Allah SWT akan selalu memberi rahmat, taufik serta hidayah-Nya kepada mereka semua.

Penulis menyadari sepenuhnya bahwa dalam penyusunan Tugas Akhir ini tidaklah sempurna, seperti kata pepatah tak ada gading yang tak retak begitu pula dalam penulisan ini, karena kesalahan dan kekurangan datangnya dari penulis sendiri serta kelebihan dan kesempurnaan datangnya dari Allah SWT. Oleh karena itu, kritik dan saran yang bersifat membangun dari para dosen, mahasiswa, dan pembaca akan sangat diharapkan dan diterima dengan senang hati. Semoga Tugas Akhir ini bermanfaat bagi semua pihak. Amin amin ya robbal ‘alamiin.

Wassalamu’alaikum, Wr.Wb

Yogyakarta, 15 Mei 2018

Leni Anggraini Susanti

DAFTAR ISI

HALAMAN JUDUL ..... **ii**
HALAMAN PERSETUJUAN PEMBIMBING ..... **iii**
HALAMAN PENGESAHAN ..... **iv**
KATA PENGANTAR ..... **v**
DAFTAR ISI ..... **vii**
DAFTAR TABEL ..... **ix**
DAFTAR GAMBAR ..... **x**
DAFTAR LAMPIRAN ..... **xi**
PERNYATAAN ..... **xii**
INTISARI ..... **xiii**
ABSTRACT ..... **xiv**
BAB I PENDAHULUAN **1**

1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Jenis Penelitian dan Metode Penelitian 4
1.5 Tujuan 4
1.6 Manfaat Penelitian 5

BAB II TINJAUAN PUSTAKA **6**

2.1 Penelitian tentang Bioinformatika 6
2.2 Penelitia tenatng *Support Vector Machine* 7
2.3 Penelitian tentang Zea Mayz L (Jagung) 9

BAB III LANDASAN TEORI **10**

3.1 Jagung (Zea Mays) 10
3.2 Bioinformatika 11

3.3 *DNA & Gene Expression* 12
3.4 *Machine Learning* 13
3.5 *Filtering* 15
3.6 *T-Test* 15
3.7 Klasifikasi 16
3.8 Klasifikasi *Support Vector Machines* 17
3.8.1 SVM Pada Data Terpisah Secara Linear 18
3.8.2 SVM Pada Data Tidak Terpisah Secara Linear 20
3.9 *Confusion Matrix* 20
3.10 *Receiver Operating Characteristic Curve* 21
BAB IV METODOLOGI PENELITIAN 23
4.1 Jenis dan Sumber Data 23
4.2 Tempat dan Waktu Penelitian 23
4.3 Variabel Penelitian 23
4.4 Metode Analisis Data 23
4.5 Langkah Penelitian 24
BAB V HASIL DAN PEMBAHASAN 25
5.1 Analisis Deskriptif 25
5.2 *Filtering* Data 26
5.3 Klasifikasi SVM 26
5.4 Hasil Klasifikasi *Support Vector Machines* (SVM) 27
5.4.1 *Confusion Matrix* 28
5.4.2 Evaluasi Model Klasifikasi 28
5.4.3 *ROC Curve* 29
5.5 Model Klasifikasi SVM 31
BAB VI KESIMPULAN DAN SARAN 32
6.1 Kesimpulan 32
6.2 Saran 33
DAFTAR PUSTAKA 34
LAMPIRAN 38

DAFTAR TABEL

Tabel 3.1 Bentuk *Confusion Matrix* dari 2 Kelas 21

Tabel 3.2 Klasifikasi AUC 22

Tabel 5.1 Data *Set Maize Growth Stages* 25

Tabel 5.2 Hasil Klasifikasi Data Latih 28

Tabel 5.3 Hasil Klasifikasi Data Uji 28

Tabel 5.4 Bobot *Gene Names* 31

DAFTAR GAMBAR

Gambar 3.1 Ikatan Hidrogen Antar nukleotida 13

Gambar 3.2 Ilustrasi SVM menemukan *hyperplane* terbaik yang memisahkan dua kelas -1 dan +1 18

Gambar 4.1 *Flowchart* Penelitian 24

Gambar 5.1 *Stripchat* Ekspresi Gen Zm00001d027230 25

Gambar 5.2 Kurva *Receiver Operating Characteristic* (ROC) Model Algoritma *Support Vector Machine* (SVM) 30

DAFTAR LAMPIRAN

Lampiran 1 : *Script Input Data* 38

Lampiran 2 : *Script* Data yang digunakan 38

Lampiran 3 : *Script* untuk memanggil data nama *gen maize*, letak *crhomosom* dan *start* 38

Lampiran 4 : *Script vector* klasifikasi 39

Lampiran 5 : *Script Filtering Data* 39

Lampiran 6 : *Script* Analisis SVM 40

Lampiran 7 : *Script* Penyusunan *Data set* 40

Lampiran 8 : *Script Tuning* Data 41

Lampiran 9 : *Script Confusion Matrix* 41

Lampiran 10 : *Script AUC* 42

Lampiran 11 : *Script* Bobot 42

Lampiran 12 : *Script stripchart gene names 1* 43

Lampiran 13 : *Script Session Info* 43

PERNYATAAN

Dengan ini penulis menyatakan bahwa dalam tugas akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepengetahuan penulis juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 19 Juli 2018



Leni Anggraini Susanti

Aplikasi *Support Vector Machines* (SVM)
Pada Data Bioinformatika

Leni Anggraini Susanti

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

INTISARI

Jagung (Zea mays L) memiliki peran dalam pemenuhan kebutuhan pangan dan strategis dibidang ekonomi yang merupakan salah satu komoditas pertanian utama. Peningkatan produktivitas jagung dapat meningkatkan ketersediaan pasokan jagung, hal ini dapat dilakukan dengan teknik budidaya yang tepat yaitu mensinergikan kebutuhan tanaman dalam setiap fase pertumbuhan dengan input yang harus diberikan. Tahap pertumbuhan jagung terdiri dari tahap pertumbuhan vegetatif dan tahap pertumbuhan reproduktif. Berdasarkan tahap pertumbuhan jagung, peneliti ingin membuat model klasifikasi dari tahap pertumbuhan jagung berdasarkan gen name pada setiap organ jaringan jagung. Maize growth stages (MGS) merupakan data pengelompokkan organ jaringan jagung berdasarkan nilai gen name apakah termasuk kelompok tahap pertumbuhan jagung vegetatif atau tahap pertumbuhan reproduktif. Dalam hal ini, digunakan sebuah cabang ilmu machine learning yang dapat mengklasifikasikan yaitu support vector machines (SVM) dengan bantuan program RStudio dengan tambahan packages bioinformatika yaitu multtest. Terdapat 3 tahap dalam pengolahannya yaitu input data, filtering untuk mereduksi data dan analisis SVM. Berdasarkan hasil analisis SVM mampu memprediksi kelas tahap pertumbuhan dengan akurasi 92.85% dengan nilai AUC 83.33%. Hal ini menunjukkan bahwa dalam penelitian ini, model algoritma Support Vector Machine (SVM) mencapai klasifikasi yang baik.

Kata Kunci : *MGS*, Bioinformatika, Klasifikasi, *Support Vector Machines*.

Aplikasi *Support Vector Machines* (SVM)
Pada Data Bioinformatika

Leni Anggraini Susanti

Departement of Statistics, Faculty of Mathematics and Natural Sciences Islamic Universit of Indonesia

ABSTRACT

Corn (Zea Mays L) has a role in the fulfillment of food and strategic needs in the field of economy which is one of the main agricultural commodities. Increased corn productivity can increase corn supply availability, this can be done with proper cultivation technique is to synergize the needs of plants in each growth phase with inputs that must be given. The growth stage of maize consists of vegetative growth stage and reproductive growth stage. Based on the growth stage of maize, the researchers wanted to make a classification model of the growth stage of maize based on the gene name on each organs of maize tissue. Maize growth stages (MGS) is a data grouping of maize tissue organs based on genes name value whether including vegetative growth stage or reproductive growth stage. In this case, used a branch of machine learning that can be classifying them with support vector machines (SVM) by the support RStudio program with bioinformatics packages multtest. There are 3 stages in the processing that includes of data input, filtering to reduce data and SVM analysis. Based on the result SVM analysis was able to predict the class growth stage with accuracy 92.85% with AUC value 83.33%. It shows that in this research, the Support Vector Machine (SVM) algorithm model achieves a good classification

Keyword : MGS, *Bioinformatics, Classifiction, Support Vector Machines*