APLIKASI METODE KLASIFIKASI K-*NEAREST NEIGHBOR* (KNN) PADA DATA *GENE EXPRESSION MAIZE Nested Association Mapping*

(Studi Kasus : Klasifikasi Organ *MAIZE Nested Association Mapping (NAM)*)

TUGAS AKHIR

Disusun Oleh:

Ika Fitria Widiawati

14 611 193

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2018

APLIKASI METODE KLASIFIKASI K-*NEAREST NEIGHBOR* (KNN) PADA DATA *GENE EXPRESSION MAIZE Nested Associaton Mapping*

(Studi Kasus : Klasifikasi Organ *MAIZE Nested Association Mapping (NAM*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana

Jurusan Statistika

Ika Fitria Widiawati

14 611 193

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2018

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Aplikasi Metode Klasifikasi K-*Nearest Neighbor* (KNN) pada Data *Gene Expression* MAIZE Nested Association Mapping (Studi Kasus : Klasifikasi Organ *Maize Nested Association Mapping (NAM)*)

Nama Mahasiswa : Ika Fitria Widiawati

Nomor Mahasiswa : 14611193

TUGAS AKHIR INI TELAH DIPERIKSAN DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, 1 Agustus 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 METODE KLASIFIKASI K-*NEAREST NEIGHBOR* (KNN) PADA DATA *GENE EXPRESSION MAIZE Nested Association Mapping*

Nama Mahasiswa : Ika Fitria Widiawati

Nomor Mahasiswa : 14 611 193

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL 1 AGUSTUS 2018

Nama Penguji	**Tanda Tangan**
1. Muryanto, S.P., M.Sc	.....
2. Arum Handini Primandari, S.Pd.Si., 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

Assalamu'alaikum Wr. Wb.

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang. Puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya kepada, shalawat serta salam kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "Aplikasi Metode Klasifikasi K-*Nearest Neighbor* (KNN) pada Data *Gene Expression* Maize *Nested Association Mapping* (Studi Kasus : Klasifikasi Organ MAIZE *Nested Association Mapping* (NAM))" dengan baik.

Tugas akhir ini disusun sebagai salah satu persyaratan yang harus dipenuhi penulis dalam menyelesaikan Program Strata Satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Tugas akhir ini penulis dengan maksimal dan memperoleh bantuan dari berbagai pihak, baik berupa bimbingan, kritik, ataupun saran. Untuk itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si selaku Ketua Jurusan Statistika.
3. Husna Nugrahapraja, Ph.D, dan Dr. techn. Rohmatul Fajriyah, M.Si selaku dosen pembimbing tugas akhir atas bimbingan dan kesabarannya dalam membimbing selama menyelesaikan tugas akhir ini.
4. Seluruh dosen Jurusan Statistika di Universitas Islam Indonesia yang telah memberikan ilmunya selama delapan semester.
5. Orang tua yang selalu setia mendoakan, memberikan dukungan, menemani serta memberikan motivasi kepada penulis selama menyelesaikan tugas akhir ini.
6. Adik kandung yang selalu mendoakan, memberikan dukungan, serta memberikan motivasi kepada penulis selama menyelesaikan tugas akhir ini.

7. Terima kasih untuk teman-teman bimbingan TA : Lina, Hawila, Ummi, Himelda, Nanda, Husain, Arie, Maidah, Leni, Shodiq, Gilang, dan Aat yang sudah sama-sama berjuan, saling mengingatkan, dan selalu menjadi sumber motivasi bagi penulis serta dorongan untuk menyelesaikan tugas akhir ini.
8. Pihak lain yang tidak dapat disebutkan satu per satu, terima kasih atas dukungan dan motivasi yang telah diberikan selama menyelesaikan tugas akhir ini.

Semoga Allah SWT selalu memberikan rahmat dan anugerah-Nya kepada mereka semua atas segala bantuan, bimbingan, motivasi dan pengajaran yang telah diberikan kepada penulis. Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu penulis menerima kritik dan saran yang bersifat membangun dari semua pihak demi kesempurnaan tugas akhir ini. Semoga tugas akhir ini bermanfaat bagi semua pihak.

Wassalamu'alaikum Wr.Wb.

Yogyakarta, 31 Juli 2018

Ika Fitria Widiawati

DAFTAR ISI

HALAMAN JUDUL ..... **i**
HALAMAN PERSETUJUAN PEMBIMBING ..... **ii**
HALAMAN PENGESAHAN ..... **iii**
HALAMAN PERSEMBAHAN ..... **iv**
KATA PENGANTAR ..... **v**
DAFTAR ISI ..... **vii**
DAFTAR TABEL ..... **viii**
DAFTAR GAMBAR ..... **ix**
DAFTAR LAMPIRAN ..... **x**
PERNYATAAN ..... **xii**
INTISARI ..... **xiii**
ABSTRACT **14**
BAB I PENDAHULUAN **1**
1.1. Latar Belakang Penelitian **1**
1.2. Rumusan Masalah **4**
1.3. Batasan Masalah **4**
1.4. Tujuan Penelitian **4**
1.5. Manfaat Penelitian **5**
BAB II TINJAUAN PUSTAKA **6**
2.1. Penelitian tentang metode *K-Nearest Neighbor (KNN)* 1
2.2. Penelitian tentang metode K-*Nearest Neighbor* pada data *Bioinformatics* **4**
2.3. Penelitian tentang *Maize* (Jagung) **4**
BAB III LANDASAN TEORI **12**
3.1. *Bioinformatics* **9**
3.2. NGS (*Next Generation Sequence*) **14**
3.3. *Gene Expression* **16**

3.4. *Filtering* **17**
3.5. *Feature Selection* **17**
3.5.1 Pengertian *Feature Selection* **17**
3.5.2 Teknik *Feature Selection* **18**
3.6. Jagung **18**
3.7. *Machine Learning* **21**
3.8. Metode Klasifikasi **21**
3.9. K-*Nearest Neighbor* **22**
3.10. *Confusion Matrix* **24**
3.11. *Kappa Statistics* **25**
3.12. *Receiver Operating Character* (ROC) **25**
BAB IV METODOLOGI PENELITIAN **27**
4.1. Jenis dan Sumber Data **27**
4.2. Variabel dan Definisi Operasional Variabel **27**
4.3. Metode Analisis Data **27**
4.4. Proses Analisis Data **28**
BAB V HASIL DAN PEMBAHASAN **30**
5.1. Deskriptif Data **30**
5.2. Proses Data *Gene Expression* **31**
5.3. Pemilihan nilai *k* **32**
5.4. Hasil Klasifikasi **34**
BAB VI PENUTUP **39**
6.1. Kesimpulan **39**
6.2. Saran **40**
DAFTAR PUSTAKA **41**
LAMPIRAN 1 ..... **44**
LAMPIRAN 2 ..... **49**
LAMPIRAN 3 ..... **53**

DAFTAR TABEL

Tabel 3.1 *Confusion Matrix* 25
Tabel 3.2 Skala Nilai *Kappa* 26
Tabel 3.3 Kualitas Klasifikasi Berdasarkan AUC 28
Tabel 5.1 *Gene Expression* Data *Maize_NAM* 32
Tabel 5.2 Tabel Nilai Akurasi dan Kappa 34
Tabel 5.3 Hasil Prediksi Klasifikasi K-NN 35
Tabel 5.4 *Confusion Matrix* 35
Tabel 5.5 *Validation* dengan *Confusion Matrix* 36
Tabel 5.6 Variabel yang Berpengaruh 38

DAFTAR GAMBAR

Gambar 3.9.1 Visualisasi Langkah Klasifikasi KNN 23
Gambar 4.1 Flowchart Penelitian..... 29
Gambar 4.1 *Flowchart* Analisis Data..... 30
Gambar 5.1 Jumlah *Gene* pada *Chromosome* 30
Gambar 5.2 Rata-rata Nilai Ekspresi Gen..... 31
Gambar 5.3 *Stripchart* Zm00001d027230 *Expression Value* 31
Gambar 5.4 Pemilihan nilai *k* terbaik 34
Gambar 5.5 Variabel *Importance*..... 38

DAFTAR LAMPIRAN

Lampiran 1:*Script* Statistik Diskriptif Data *Maize NAM* 47
Lampiran 2: *Script Filtering* dan *Feature Selection* 48
Lampiran 3: *Script* Analisis KNN dan Hasil Akurasi 51

DAFTAR ISTILAH

NAM	: *Nested Association Mapping*
KNN	: *K-Nearest Neighbor*
NGS	: *Next generation sequencing*
ROC	: *Receiver Operating Characteristic*
AUC	: *Area Under Curve*
DNA	: *Deoxyribo Nucleic Acid*

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya 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, 31 Juli 2018

Ika Fitria Widiawati

APLIKASI METODE KLASIFIKASI K-*NEAREST NEIGHBOR* (KNN) PADA DATA *GENE EXPRESSION MAIZE Nested Associaton Mapping*

(Studi Kasus : Klasifikasi Organ *MAIZE Nested Association Mapping (NAM*)

Oleh : Ika Fitria Widiawati

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

INTISARI

Jagung merupakan tanaman yang memiliki nilai ekonomi tinggi. Salah satu cara yang digunakan untuk meningkatkan produksi jagung yaitu dengan menerapkan teknik Nested Association Mapping (NAM). Teknik NAM dirancang untuk mengidentifikasi dan membedah arsitektur genetik ciri-ciri kompleks dalam jagung (Zea mays). Data ekspresi gen dalam jumlah yang besar memberikan tantangan dalam mengaplikasikan metode analisis, menginterpretasi, dan mengambil kesimpulan. Penelitian ini bertujuan mengevaluasi hasil klasifikasi gene expression pada varietas jagung mengenai klasifikasi organ-organ nya, yaitu Apex, Root, Shoot, Ear, dan Tassel dengan menggunakan K-Nearest Neighbor (KNN). Berdasarkan analisis KNN dengan k=5 diperoleh nilai akurasi dan nilai AUC masing masing sebesar 0. 926 dan sebesar 0.992. Hal ini menunjukkan klasifikasi gene expressions pada varietas jagung untuk membedakan organ Apex, root, shoot, ear, dan tassel dapat dipercaya.

Kata Kunci: Klasifikasi, *gene expressions*, Jagung, *K- Nearest Neighbor (KNN*), Kappa

APPLICATION OF K-NEAREST NEIGHBOR (KNN) CLASSIFICATION ON GENE EXPRESSION DATA

(Case Study: Organ Classification on MAIZE Nested Association Mapping (NAM))

By : Ika Fitria Widiawati

ABSTRACT

Maize is a plant that has high economic value. One method used to increase corn production is by applying the Nested Association Mapping (NAM) technique. NAM techniques are designed to identify and dissect the genetic architecture of complex features in maize (Zea mays). Genes expression data in large numbers provides challenges in applying analytical methods, interpreting, and drawing conclusions. This study aim to evaluate the results of gene expression classification on corn varieties regarding the classification of its organs, namely Apex, Root, Shoot, Ear and Tassel by using K-Nearest Neighbor (KNN). Based on KNN analysis with k = 5, we obtained the accuracy value and AUC value of 0. 926 and 0.992 respectively and showed the classification on maize lines to distinguish Apex, root, shoot, ear, and tassel organs based on the genes expression.

Keywords: Classification, gene expressions, Maize, K- Nearest Neighbor (KNN), Kappa