APLIKASI KLASIFIKASI *K-NEAREST NEIGHBOR* (K-NN) PADA PASIEN KANKER PARU-PARU

(Studi Kasus : Klasifikasi Stadium Pasien Kanker Paru-paru Berdasarkan *Gene Expression Data*di Amerika Serikat, GSE10072)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

Jurusan Statistika

Disusun oleh :

Ines Riantika

13611192

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2018

HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR

Judul	:	Aplikasi Klasifikasi *K-Nearest Neighbor* (K-NN) pada Kanker Paru-paru (Studi Kasus : Klasifikasi Stadium Pasien Kanker Paru-paru Berdasarkan *Gene Expression* Data di Amerika, GSE10072)
Nama Mahasiswa	:	Ines Riantika
Nomor Mahasiswa	:	13611192

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN
Yogyakarta, 03 Januari 2018

Pembimbing

(Dr.techn. Rohmatul Fajriyah, S.Si, M.Si)

i

HALAMAN PENGESAHAN

TUGAS AKHIR

APLIKASI KLASIFIKASI *K-NEAREST NEIGHBOR (K-NN)*
PADA PASIEN KANKER PARU-PARU

(Studi Kasus : Klasifikasi Stadium Pasien Kanker Paru-paru Berdasarkan *Gene Expression Data* di Amerika Serikat, GSE10072)

Nama Mahasiswa : Ines Riantika

Nomor Mahasiswa : 13611192

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 13 FEBRUARI 2018

Nama Penguji	**Tanda Tangan**
1. Fitria Dyah Ayu Suryanegara, M.Sc., Apt	.....
2. Ayundyah Kesumawati, M.Si	.....
3. Dr.techn.Rohmatul Fajriyah, S.Si, M.Si	.....

Mengetahui,
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

(Drs. Allwar., M.Sc., P.hD)

ii

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Assalamu'alaikum, wr, wb.

Alhamdulillahirabbil'alamin, Puji Syukur saya panjatkan kehadirat Allah SWT atas berkat rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan penulisan skripsi ini yang berjudul **"Aplikasi Klasifikasi *K-Nearest Neighbor* (K-NN)Pada Pasien Kanker Paru-paru**". Untuk memperoleh gelar sarjana di Jurusan Statistika dapat terselesaikan tanpa hambatan yang berarti. Shalawat serta salam semoga selalu tercurah kepada Nabi Muhammad SAW.

Pada kesempatan ini penulis mengucapkan terimakasih kepada dosen pembimbing penulis Dr.tench.RohmatulFajriyah.S.Si.,M.Si. yang selalu memberikan dukungan, dan saran dalam menyelesaikan penelitian ini. Trimakasih penulis berikan terhadap pihak-pihak lainya :

1. Bapak Drs. Allwar, M.Sc.,Ph.D, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia
2. Bapak Dr. RB Fajriya Hakim, M.Si., selaku ketua prodi statistika, beserta jajarannya.
3. Kedua Orang tua tercinta, Ahmad Kamildan ibu Maisarah serta Kakak-kakak dan saya, Mutia Elvarina, Azlan Romil, Fikri Aldi dan Afifah Azmi yang selalu memberikan doa, dukungan, kasih sayang yang menjadi semnagat terbesar untuk menyelesaikan skripsi ini.
4. Teman seperjuangan Rizka Wulandari, Ditha Runita, Mega Ayu Lestari, dan Suci Eka Purnamasari yang memberikan semangat dan doronganya ketika mnyusun skripsi ini.
5. Seluruh teman-teman penulis Mayda Rosa Devina, Novita Kharimatunnisa, Novia Nigsih, Siti Amelia Mubarokkah, Nevy Tarigan,Yayi Diyah dan Intan Anggraini yang memberikan saran dan doronganya selama penulisan.

Semoga Allah SWT membalas segala kebaikan mereka dengan segala anugrah, rahmat, dan Hidayah-Nya. Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari sempurna, oleh karena itu segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga Tugas Akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan umumnya. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal alamiin.

Wassalamu'alaikum Wr.Wb.

Yogyakarta, 16 Februari 2017

Penulis,

(Ines Riantika)

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

PERNYATAAN

Dengan ini penulis menytakan bahwa Tugas Akhir ini tidak terdapat karya orang lain yang sebelumnya digunakan untuk meperoleh gelar kesarjanaan di suatu perguruan tinggi, dan tidak terdapat karya yang pernah diterbitkan oleh orang lain, kecuali semua yang dijadikan referensi yang telah disebutkan dalam daftar pustaka.

Yogyakarta, 16 Februari 2016

Penulis

APLIKASI KLASIFIKASI *K-NEAREST NEIGHBOR(K-NN)* PADA *GENE EXPRESSION MICROARRAY*

(Studi Kasus :Klasifikasi Stadium Pasien Kanker Paru-paru Berdasarkan *Gene Expression* Data Di Amerika, GSE10072)

Oleh : Ines Riantika
Program Studi Statistika Fakultas MIPA
Universitas Islam Indonesia

INTISARI

Kanker adalah suatu penyakit yang disebabkan oleh tidak terkontrolnya pertumbuhan dan penyebaran sel yang tidak normal. Kanker adalah penyebab kematian terbesar didunia,bahkan pada tahun2017 di estimasikan penderita baru kanker di Amerika Serikat mencapai 1.688.780 jiwa. Salah satu jenis kanker yang paling banyak diderita oleh warga Amerika adalah jenis kanker paru-paru dengan tingkat kematian pria pada tahun 2017 mencapai 155.870 jiwa. Stadium pada kanker biasanya dilakukan untuk melakukan bagaimana pengobatan yang disarankan untuk pasien. Maka dari itu pada penelitian tersebut dilakukan klasifikasi menggunakan data *gene expression* untuk mengklasifikasikan stadium kanker paru-paru *non small cell lung cancer*(NSCLC) untuk membedakan jaringan tumor stadium awal dan jaringan tumor stadium akhir dengan menggunakan $k = 9$ didapatkandidapat hasil akurasi dari *confusion matrix* sebesar 0.848, dan ROC dengan nilai AUC 0.9306 yang juga diartikan bahwa klasifikasi baik untuk data ini.

Kata kunci : Klasifikasi , *Gene expression,* Kanker, K-*nearest neighbor*

APPLICATIONCLASSIFICATIONOF K-NEAREST NEIGHBOR (K-NN) ON GENE EXPRESION *MICROARRAY*

(Case Study: Classification of Stage Patients of Lung Cancer In America Using Gene Expression Data GSE10072)

Studi Kasus : Klasifikasi Stadium Pasien Kanker Paru-paru Berdasarkan *Gene Expression* Data Di Amerika, GSE10072

By : Ines Riantika

Supervised by : Dr. Tench. Rohmatul Fajriyah, S.Si., M.Si.

ABSTRACT

Cancer is a group of diseases characterized by the uncontrolled growth and spread of abnormal cell. Cencer is the biggest cause of death in the world. Estimated numbers of new cancer cases for 2017 at United States reached 1,688,780 people. One type cancer with hight death rate in 2017 is lung cncer with 155.870 people. Staging in cancer is usually for given the recommended treatment for the patient. Therefore, in this study, based of gene expression data for classify the stages of lung cancer of non-small cell lung cancer (NSCLC) for early-stage tumor tissue and late-stage tumor tissue using k = 9. it is found confusion matrix with 0.848, and ROC with AUC value 0.9306 which is also interpreted good K-NN *classification.*

Keyword : Classifiaction , Gene expression, Cancer, K-nearest neighbor