

**PENERAPAN ALGORITMA *MODIFIED* K-NEAREST
NEIGHBOR (*MKNN*) UNTUK MENGLASIFIKASIKAN
LETAK PROTEIN PADA BAKTERI *E-COLI***

SKRIPSI

Oleh:

SULISTYANDARI

0610963060-96



**PROGRAM STUDI ILMU KOMPUTER
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2012**

**PENERAPAN ALGORITMA *MODIFIED* K-NEAREST
NEIGHBOR (*MKNN*) UNTUK MENGLASIFIKASIKAN
LETAK PROTEIN PADA BAKTERI *E-COLI***

SKRIPSI

Sebagai salah satu syarat untuk memperoleh
gelar Sarjana Komputer dalam bidang Ilmu Komputer

Oleh :

Sulistiyandari
0610963060-96



**PROGRAM STUDI ILMU KOMPUTER
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2012**

LEMBAR PENGESAHAN SKRIPSI

**PENERAPAN ALGORITMA *MODIFIED* K-NEAREST
NEIGHBOR (*MKNN*) UNTUK MENGLASIFIKASIKAN
LETAK PROTEIN PADA BAKTERI *E-COLI***

Oleh:
SULISTYANDARI
0610963060-96

Setelah dipertahankan di depan Majelis Penguji
Pada tanggal 1 Februari 2012
dan dinyatakan memenuhi syarat untuk memperoleh
gelar Sarjana Komputer dalam bidang Ilmu Komputer

Pembimbing I

Lailil Muflikhah, S.Kom, MSc
NIP. 197411132005012001

Pembimbing II

Yusi Tyroni .M,S.Kom,MS
NIP. 198002282006041001

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr. Abdul Rouf Alghofari, M.Sc.
NIP. 196709071992031001

UNIVERSITAS BRAWIJAYA



LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Sulistyandari
NIM : 0610963060-96
Jurusan : Matematika
Program Studi : Ilmu Komputer

Penulis Skripsi berjudul : Penerapan Algoritma Modified K-Nearest Neighbor (*MKNN*) untuk mengklasifikasikan letak protein pada bakteri E-Coli.

Dengan ini menyatakan bahwa :

1. Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub dalam isi dan tertulis pada daftar pustaka dalam Skripsi ini.
2. Apabila dikemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala resiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 1 Februari 2012
Yang menyatakan,

(Sulistyandari)
NIM. 0610963060-96

UNIVERSITAS BRAWIJAYA



PENERAPAN ALGORITMA *MODIFIED K-NEAREST NEIGHBOR (MKNN)* UNTUK MENGLASIFIKASIKAN LETAK PROTEIN PADA BAKTERI *E-COLI*

ABSTRAK

Bakteri *E.Coli* merupakan jenis bakteri patogen yang potensial. Manusia dapat diuntungkan ataupun dirugikan tergantung dari potensi yang diberikan oleh bakteri tersebut. Fungsi potensial bakteri *E.Coli* dipengaruhi oleh penyebaran protein pada selnya, sehingga diperlukan tehnik klasifikasi agar dapat diketahui jenis *E.Coli* mana saja yang dapat merugikan berdasarkan tata letak protein pada selnya.

Algoritma *MODIFIED K-NEAREST NEIGHBOR (MKNN)* merupakan sebuah algoritma yang digunakan untuk melakukan klasifikasi. *MKNN* berdasar pada label kelas data sesuai dengan parameter k yang ditetapkan dengan perhitungan *Weighted k-NEAREST NEIGHBOR*.

Data yang digunakan adalah 336 dengan 7 atribut (bertipe kontinue). Nilai akurasi terbaik untuk dataset *E.Coli* pada record ke 300 rasio ke 6:4 dengan akurasi maksimum 78,33% dan akurasi minimumnya pada record ke 50 dengan akurasi 20%.

Kata kunci : Bakteri *E.Coli*, dataset, *MKNN*.

UNIVERSITAS BRAWIJAYA



IMPLEMENTATION OF MODIFIED K-NEAREST NEIGHBOR (MKNN) ALGORITHM TO CLASSIFY PROTEIN LAYOUT OF E-COLI BACTERIA

ABSTRACT

E. Coli bacteria is a type of potentially pathogenic bacteria. Humans can be benefited or harmed depending on the potential offered by these bacteria. Potential function of E. Coli is affected by the spread of the proteins within the cel, so classification techniques are necessary in order to acknowledge which type of E. Coli that can harm based on the protein layout of the cell.

Modified k-Nearest Neighbor is an algorithm used to perform classification. MKNN class label based on the data according to the parameter k which is determined by the calculation of Weighted k-Nearest Neighbor.

The data used was 336 with 7 attributes (continuous type). The best accuracy value for the E. Coli dataset is 78.33% as maximum accuracy on the record of 300 with 6:4 ratio and minimum accuracy of 20% when the record is 50.

Keywords : E. Coli bacteria, dataset, MKNN

UNIVERSITAS BRAWIJAYA



Kata Pengantar

Alhamdulillah rabbil 'alamin. Puji syukur penulis panjatkan kehadiran Allah SWT, atas segala rahmat dan limpahan hidayah-Nya tugas akhir yang berjudul “*PENERAPAN ALGORITMA MODIFIED K-NEAREST NEIGHBOR (MKNN) UNTUK MENGLASIFIKASIKAN LETAK PROTEIN PADA BAKTERI E.COLI*” ini dapat terselesaikan. Skripsi ini disusun dan diajukan sebagai syarat untuk memperoleh gelar sarjana pada program studi Ilmu Komputer, Jurusan Matematika, Fakultas MIPA, Universitas Brawijaya

Semoga Allah melimpahkan rahmat atas Nabi Muhammad SAW, makhluk paling mulia yang senantiasa memberikan cahaya petunjuk, dan atas keluarganya dan sahabat-sahabatnya.

Dalam penyelesaian tugas akhir ini, penulis mendapat banyak bantuan baik moral maupun materiil dari banyak pihak. Atas bantuan yang telah diberikan, penulis menyampaikan penghargaan dan ucapan terima kasih yang sedalam-dalamnya kepada:

1. Lailil Muflikhah., S.Kom., M.Sc, dan Yusi Tyroni.M, S.Kom., MS., selaku dosen pembimbing, terima kasih atas semua saran, bantuan, kritikan, waktu, dorongan semangat dan bimbingannya.
2. Drs. Marji, MT., selaku Ketua Program Studi Ilmu Komputer Universitas Brawijaya Malang.
3. Edy Santoso S.Si., M.Kom., selaku dosen Penasihat Akademik.
4. Segenap bapak dan ibu dosen yang telah mendidik dan mengamalkan ilmunya kepada penulis.
5. Segenap staf dan karyawan di Jurusan Matematika FMIPA Universitas Brawijaya.
6. Kedua orang tua serta semua keluarga besarku terima kasih atas semua cinta, kasih sayang, doa, dukungan, nasehat, dan semangat yang tiada henti.
7. Teman-teman ilkomers 2006, khususnya ilkomers B 2006 ‘Babi Inside’ terima kasih untuk semangat, cinta, dukungan, dan kebersamaan yang diberikan.
8. Teman-teman kos JL Bunga Andong 8A terima kasih untuk kebersamaan selama ini.

9. Falentina Nugraheni.P, Duratun Nafisah, Riris .F dan Tiara.A yang selalu bersedia membantu memberikan ide dalam penyelesaian skripsi saya ini.
10. Pihak lain yang telah membantu terselesaikannya Tugas Akhir ini yang tidak bisa penulis sebutkan satu-persatu.

Semoga penulisan laporan tugas akhir ini bermanfaat bagi pembaca sekalian. Sebagai manusia, penulis menyadari tugas akhir ini masih jauh dari kesempurnaan dan banyak kekurangan, dengan kerendahan hati penulis mengharapkan kritik dan saran yang membangun dari pembaca.

Malang, 01 Februari 2012



SULISTYANDARI

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
LEMBAR PERNYATAAN.....	v
ABSTRAK.....	vii
ABSTRACT.....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
DAFTAR KODE PROGRAM.....	xix
DAFTAR LAMPIRAN.....	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metode Penyelesaian Masalah.....	3
1.7 Sistematika Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	7
2.1 Algoritma.....	7
2.2.1 Definisi Algoritma.....	7
2.2 Data.....	7
2.2.1 Definisi Data.....	7
2.3 Data Mining.....	9
2.3.1 Definisi <i>Data Mining</i>	9
2.3.2 Proses <i>Data Mining</i>	11
2.3.3 Metode <i>Data Mining</i>	11
2.4 Klasifikasi.....	11
2.5 <i>K-Nearest Neighbor (KNN)</i>	12
2.5.1 Definisi <i>K-Nearest Neighbor (KNN)</i>	12
2.5.2 Proses <i>K-Nearest Neighbor (KNN)</i>	13
2.6 <i>Modified K-Nearest Neighbor (MKNN)</i>	14
2.6.1 Definisi <i>Modified K-Nearest Neighbor (MKNN)</i>	14

2.6.2	Proses <i>Modified K-Nearest Neighbor (MKNN)</i>	14
2.7	Bakteri <i>E. Coli</i>	16
2.8	Perhitungan Akurasi	19
BAB III METODOLOGI DAN PERANCANGAN		21
3.1	Analisa Sistem	22
3.1.1	Deskripsi Sistem	22
3.1.2	Deskripsi Data	22
3.2	Perancangan Proses	23
3.2.1	Proses Klasifikasi Algoritma <i>MKNN</i>	23
3.2.1.1	<i>Flowchart</i> Sistem	24
3.2.1.2	Proses Klasifikasi <i>MKNN</i>	25
3.2.1.3	Menghitung Validitas	26
3.2.1.4	Menghitung Euclidean	27
3.2.1.5	Menghitung Weight Voting	29
3.3	Contoh Perhitungan Manual	30
3.3.1	Data <i>Testing</i> dan Data <i>Traning</i> pada data letak protein pada bakteri <i>E.Coli</i>	30
3.4	Perancangan Antarmuka	33
3.5	Perancangan Uji Coba	35
3.5.1	Menentukan Akurasi Data	35
BAB IV IMPLEMENTASI DAN PEMBAHASAN		37
4.1	Perangkat Sistem	37
4.1.1	Perangkat Lunak	37
4.1.2	Perangkat Keras	37
4.2	Implementasi	37
4.2.1	Implementasi Program	38
4.2.1.1	Proses Input	38
4.2.1.1.1	Input file data	38
4.2.1.1.2	Membaca dan memasukkan data pada table	39
4.2.1.1.3	Mengurutkan kelas pada data bakteri dari atas	41
4.2.1.1.3	Mengurutkan kelas pada data bakteri dari bawah	42
4.2.1.2	Proses Klasifikasi <i>MKNN</i>	42
4.2.1.2.1	Hitung Validitas	45
4.2.1.2.2	Hitung Euclidean	45

4.2.1.2.3	Hitung Nilai Weight Voting	46
4.3	Penerapan Aplikasi.....	46
4.3.1	Implementasi Antarmuka	46
4.3.1.1	Tabpage 1.....	46
4.3.1.2	Tabpage 2.....	48
4.4	Implementasi Pengujian.....	49
4.4.1	Hasil Uji	49
4.4.2	Tabel Pengujian pada data set E.Coli.....	49
4.4.2.1	Pengujian Pada Record 50.....	50
4.4.2.2	Pengujian Pada Record 100.....	51
4.4.2.3	Pengujian Pada Record 200.....	52
4.4.2.4	Pengujian Pada Record 300.....	53
4.5	Analisa Hasil	54
4.5.1	Pengujian untuk mengetahui pengaruh nilai k terhadap tingkat akurasi.....	55
4.5.1.1	Hasil pengujian untuk mengetahui pengaruh nilai k terhadap tingkat akurasi.....	55
4.5.2	Pengujian untuk mengetahui pengaruh jumlah data latih (record) terhadap tingkat akurasi	56
4.5.2.1	Hasil pengujian untuk mengetahui pengaruh jumlah data latih (record) terhadap tingkat akurasi	57
BAB V KESIMPULAN DAN SARAN.....		59
5.1	Kesimpulan.....	59
5.2	Saran.....	59
DAFTAR PUSTAKA.....		61
LAMPIRAN		63

UNIVERSITAS BRAWIJAYA



DAFTAR GAMBAR

Gambar 2.1 Proses <i>Knowledge discovery in databases</i> (KDD) (Usama Fayyad, 1996)	8
Gambar 2.2 Taksonomi Metode Data Mining	10
Gambar 2.3 Struktur Sel Bakteri E.Coli.....	17
Gambar 3.1 Alur Penelitian.....	21
Gambar 3.2 <i>Flowchart</i> Sistem Keseluruhan	24
Gambar 3.3 <i>Flowchart</i> Proses Klasifikasi MKNN.....	25
Gambar 3.4 <i>Flowchart</i> Menghitung <i>Validitas</i>	27
Gambar 3.5 <i>Flowchart</i> Menghitung <i>Euclidean</i>	28
Gambar 3.6 <i>Flowchart</i> Menghitung <i>Weight Voting</i>	29
Gambar 3.7 Tabpage 1 pada Program.....	33
Gambar 3.8 Tabpage 2 pada Program.....	34
Gambar 4.1 Tabpage 1 sebelum diproses.....	47
Gambar 4.2 Tabpage 1 setelah diproses.....	48
Gambar 4.3 Tabpage 2 Tampilan Akurasi	48
Gambar 4.4 Pengaruh nilai k terhadap tingkat akurasi	55
Gambar 4.5 Pengaruh jumlah nilai maksimum tiap record terhadap tingkat akurasi.....	56

UNIVERSITAS BRAWIJAYA



DAFTAR TABEL

Tabel 3.1 Tabel Data <i>Testing</i> dan Data <i>Training</i>	30
Tabel 3.2 Tabel Perhitungan Validitas.....	31
Tabel 3.3 Tabel Perhitungan <i>Euclidean</i>	31
Tabel 3.4 Tabel Perhitungan <i>Weight Voting</i>	32
Tabel 3.5 Rancangan Tabel Pengujian Akurasi.....	34
Tabel 4.1 Tabel pengujian pada data set <i>E. Coli</i> untuk record 50	50
Tabel 4.2 Tabel pengujian pada data set <i>E. Coli</i> untuk record 100	51
Tabel 4.3 Tabel pengujian pada data set <i>E. Coli</i> untuk record 200	53
Tabel 4.4 Tabel pengujian pada data set <i>E. Coli</i> untuk record 300	54



UNIVERSITAS BRAWIJAYA



DAFTAR KODE PROGRAM

<i>Sourcecode</i> 4.1 Input file data	38
<i>Sourcecode</i> 4.2 Membaca dan memasukkan data pada table	39
<i>Sourcecode</i> 4.3 Mengurutkan kelas pada data bakteri dari atas	41
<i>Sourcecode</i> 4.4 Mengurutkan kelas pada data bakteri dari bawah	42
<i>Sourcecode</i> 4.5 Proses klasifikasi dengan membagi data latih dan data uji	43
<i>Sourcecode</i> 4.6 Proses klasifikasi menentukan kelas pada data testingnya	44
<i>Sourcecode</i> 4.7 Hitung Validitas	45
<i>Sourcecode</i> 4.8 HitungEuclidean	46
<i>Sourcecode</i> 4.9 Weight Voting	46



UNIVERSITAS BRAWIJAYA



DAFTAR LAMPIRAN

Lampiran 1. Tabel Penelitian Bakteri.....	65
Lampiran 2. Pengaruh Data Latih.....	75
Lampiran 3. Data Bakteri E-Coli.....	75

UNIVERSITAS BRAWIJAYA



UNIVERSITAS BRAWIJAYA

