

**DIAGNOSIS TINGKAT RISIKO PENYAKIT STROKE  
MENGUNAKAN METODE *K-NEAREST NEIGHBOR* DAN  
*NAÏVE BAYES***

**SKRIPSI**

Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun oleh:

Annisa Puspitawuri

NIM: 135150201111108



**PROGRAM STUDI TEKNIK INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BRAWIJAYA  
MALANG  
2019**

**PENGESAHAN**

**DIAGNOSIS TINGKAT RISIKO PENYAKIT STROKE MENGGUNAKAN METODE  
K-NEAREST NEIGHBOR DAN NAÏVE BAYES**

**SKRIPSI**

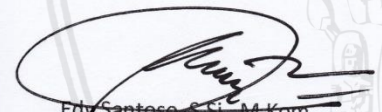
Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

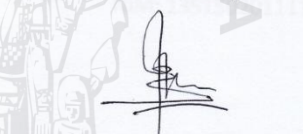
Disusun Oleh :  
Annisa Puspitawuri  
NIM: 135150201111108

Skripsi ini telah diuji dan dinyatakan lulus pada  
2 Januari 2019  
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I


Dosen Pembimbing II

  
Edy Santoso, S.Si., M.Kom.  
NIP: 19740414 200312 1 004

  
Candra Dewi, S.Kom., M.Sc.  
NIP: 19771114 200312 2 001



Mengetahui  
Ketua Jurusan Teknik Informatika

  
Tri Astoto Kurniawan, S.T., M.T., Ph.D.  
NIP: 19710518 200312 1 001

### PERNYATAAN ORISINALITAS

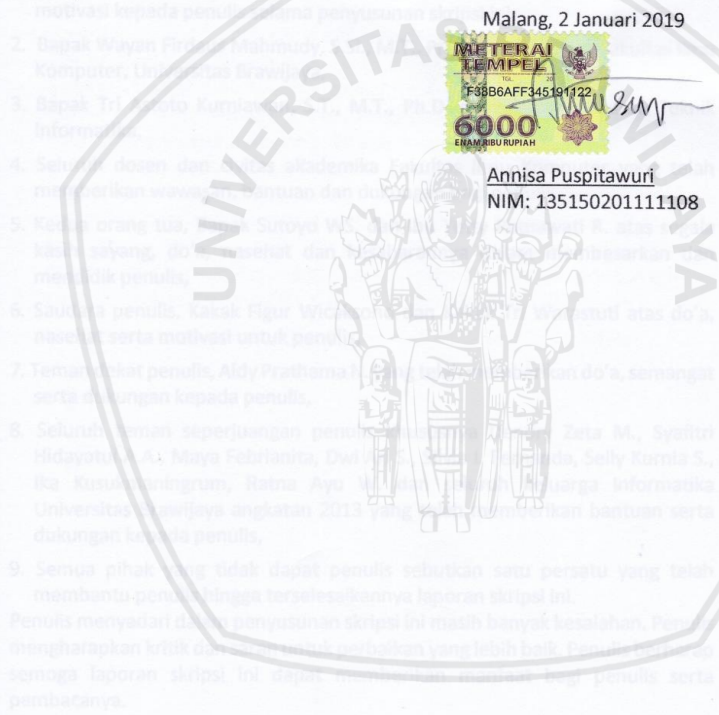
Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 2 Januari 2019



Annisa Puspitawuri  
NIM: 135150201111108



Malang, 2 Januari 2019

Penulis  
annispwtwr@gmail.com

## PRAKATA

Segala puji bagi Allah SWT yang Maha Pengasih dan Maha Penyayang, berkat limpahan rahmat serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi “Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode *K-Nearest Neighbor* dan *Naïve Bayes*” dengan baik. Skripsi ini ditujukan sebagai syarat ujian seminar skripsi dalam rangka untuk memperoleh gelar Sarjana Komputer di Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya Malang. Atas terselesaikannya skripsi ini, penulis mengucapkan terimakasih kepada:

1. Bapak Edy Santoso, S.Si., M.Kom. dan Ibu Candra Dewi, S.Kom., M.Sc. selaku Dosen Pembimbing skripsi yang telah membimbing, memberi saran serta motivasi kepada penulis selama penyusunan skripsi ini,
2. Bapak Wayan Firdaus Mahmudy, S.Si., M.T., Ph.D. selaku Dekan Fakultas Ilmu Komputer, Universitas Brawijaya,
3. Bapak Tri Astoto Kurniawan, S.T., M.T., Ph.D. selaku Ketua Jurusan Teknik Informatika,
4. Seluruh dosen dan civitas akademika Fakultas Ilmu Komputer yang telah memberikan wawasan, bantuan dan dukungan bagi penulis,
5. Kedua orang tua, Bapak Sutoyo WS. dan Ibu Yetty Ratnawati R. atas segala kasih sayang, do’a, nasehat dan kesabarannya dalam membesarkan dan mendidik penulis,
6. Saudara penulis, Kakak Figur Wicaksono dan Kakak Tri Warastuti atas do’a, nasehat serta motivasi untuk penulis,
7. Teman dekat penulis, Aldy Prathama N. yang telah memberikan do’a, semangat serta dukungan kepada penulis,
8. Seluruh teman seperjuangan penulis khususnya Dendry Zeta M., Syafitri Hidayatul A.A., Maya Febrianita, Dwi Ari S., Silvia I. Fernanda, Selly Kurnia S., Ika Kusukmaningrum, Ratna Ayu W. dan seluruh keluarga Informatika Universitas Brawijaya angkatan 2013 yang telah memberikan bantuan serta dukungan kepada penulis,
9. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu penulis hingga terselesaikannya laporan skripsi ini.

Penulis menyadari dalam penyusunan skripsi ini masih banyak kesalahan. Penulis mengharapkan kritik dan saran untuk perbaikan yang lebih baik. Penulis berharap semoga laporan skripsi ini dapat memberikan manfaat bagi penulis serta pembacanya.

Malang, 2 Januari 2019

Penulis  
annisapsptwr@gmail.com

## ABSTRAK

**Annisa Puspitawuri, Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode *K-Nearest Neighbor* dan *Naïve Bayes***

**Pembimbing: Edy Santoso, S.Si., M.Kom. dan Candra Dewi, S. Kom., M.Sc.**

Penyakit stroke merupakan penyakit yang timbul akibat terputusnya suplai darah menuju otak karena terdapat semburan pembuluh darah atau terjadi sumbatan berupa gumpalan darah. Stroke merupakan penyebab kecacatan nomor satu dan penyebab kematian nomor tiga di dunia setelah penyakit jantung dan kanker, baik di negara maju maupun berkembang. Berdasarkan data Riset Kesehatan Dasar, prevalensi stroke di Indonesia pada tahun 2013 mengalami kenaikan jika dibandingkan dengan data Riskesdas 2007 dengan nilai angka 8,3%, naik mencapai angka 12,1% per 1.000 penduduk. Untuk itu diperlukan suatu tindakan pendeteksian tingkat risiko penyakit stroke agar dapat segera diatasi sesuai dengan tingkat risikonya. Penelitian ini mengusulkan adanya suatu aplikasi diagnosis tingkat risiko penyakit stroke menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* karena data yang didapat menggunakan atribut numerik dan kategoris. Algoritme *K-Nearest Neighbor* digunakan untuk memproses data numerik, dan algoritme *Naïve Bayes* digunakan untuk memproses data kategoris. Hasil penelitian menunjukkan nilai akurasi tertinggi yang diperoleh pada data kelas seimbang adalah 96.67% dengan data latih 45, data uji 30 dan nilai  $K=15-22$ . Sedangkan pada data latih tidak seimbang, menunjukkan akurasi tertinggi sebesar 100% dengan jumlah data latih 60, data uji 30 dan nilai  $K=20-30$ .

Kata Kunci: penyakit stroke, klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes*

## ABSTRACT

**Annisa Puspitawuri, Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode *K-Nearest Neighbor* dan *Naïve Bayes***

**Supervisors: Edy Santoso, S.Si., M.Kom. dan Candra Dewi, S. Kom., M.Sc.**

*Stroke is a disease that arises due to the dissolution of blood supply to the brain because of bursts in the blood vessels or there was a blockage of blood clots. It is a number 3 cause of death after heart disease and cancer and stroke is a leading cause of disability, both in developed and developing countries. Based on Riskesdas data, stroke prevalence in Indonesia in 2013 has increased when compared with Riskesdas data in 2007 with a value of 8.3%, increase up to 12.1% per 1,000 population. Therefore, we need an action to detect the level of risk of stroke to be immediately addressed in accordance with the level of risk. This research proposes an application of diagnosis of stroke risk level using *K-Nearest Neighbor* and *Naïve Bayes* methods, because the data obtained using numerical and categorical attributes. *K-Nearest Neighbor* algorithm is used to process numerical data, and *Naïve Bayes* algorithm is used to process categorical data. The results showed that the highest accuracy value obtained in the balanced class data was 96.67% with 45 training datasets, 30 testing datasets and value of  $K=15-22$ . Meanwhile, the training datasets that is not balanced shows the highest accuracy of 100% with the number of training datasets is 60, 30 testing datasets and the value of  $K=20-30$ .*

**Keywords:** *stroke, classification, K-Nearest Neighbor, Naïve Bayes*

## DAFTAR ISI

PENGESAHAN .....	i
PERNYATAAN ORISINALITAS .....	ii
PRAKATA.....	iii
ABSTRAK.....	iv
ABSTRACT .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR KODE PROGRAM .....	xi
DAFTAR LAMPIRAN .....	xii
BAB 1 PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan .....	3
1.4 Manfaat.....	3
1.5 Batasan Masalah .....	3
1.6 Sistematika Pembahasan .....	3
BAB 2 LANDASAN KEPUSTAKAAN .....	5
2.1 Kajian Pustaka .....	5
2.2 Penyakit Stroke .....	7
2.2.1 Jenis Penyakit Stroke.....	7
2.2.2 Faktor Risiko Penyakit Stroke.....	8
2.3 <i>K-Nearest Neighbor</i> .....	9
2.4 <i>Naïve Bayes</i> .....	10
2.5 <i>K-Nearest Neighbor dan Naïve Bayes</i> .....	10
BAB 3 METODOLOGI .....	12
3.1 Kajian Pustaka .....	12
3.2 Pengumpulan Data .....	13
3.3 Perancangan .....	13
3.4 Implementasi .....	13

3.5 Pengujian dan Analisis Hasil.....	14
3.6 Kesimpulan dan Saran .....	14
BAB 4 PERANCANGAN.....	15
4.1 Perancangan Proses.....	15
4.1.1 Proses <i>K-Nearest Neighbor</i> .....	16
4.1.2 Proses <i>Naïve Bayes</i> .....	18
4.2 Deskripsi Data .....	19
4.3 Perhitungan Manual .....	21
4.3.1 Menentukan Nilai <i>K</i> .....	24
4.3.2 Menghitung Jarak <i>Euclidean</i> .....	24
4.3.3 Mengurutkan Jarak Terdekat.....	27
4.3.4 Pengambilan Data Latih Sebanyak <i>K</i> .....	28
4.3.5 Menghitung Probabilitas Prior .....	28
4.3.6 Menghitung Probabilitas Likelihood .....	29
4.3.7 Menghitung Posterior .....	29
4.4 Perancangan Antarmuka .....	30
4.4.1 Perancangan Antarmuka Tampilan Data .....	30
4.4.2 Perancangan Antarmuka Klasifikasi .....	31
4.4.3 Perancangan Antarmuka Pengujian.....	31
4.5 Perancangan Pengujian .....	32
4.5.1 Pengujian Data Kelas Seimbang dan Tidak Seimbang .....	32
BAB 5 IMPLEMENTASI .....	33
5.1 Spesifikasi Sistem .....	33
5.1.1 Spesifikasi Perangkat Keras.....	33
5.1.2 Spesifikasi Perangkat Lunak .....	33
5.2 Batasan Implementasi .....	33
5.3 Implementasi Algoritme .....	34
5.3.1 Implementasi Algoritme <i>K-Nearest Neighbor</i> .....	34
5.3.2 Implementasi Algoritme <i>Naïve Bayes</i> .....	35
5.4 Implementasi Antarmuka Tampilan Data.....	43
5.5 Implementasi Antarmuka Klasifikasi.....	43
5.6 Implementasi Antarmuka Pengujian .....	44





BAB 6 PENGUJIAN DAN ANALISIS.....	45
6.1 Pengujian Sebaran Data Kelas Seimbang .....	45
6.2 Pengujian Sebaran Data Kelas Tidak Seimbang.....	46
6.3 Analisis Hasil Pengujian.....	47
BAB 7 PENUTUP .....	50
7.1 Kesimpulan.....	50
7.2 Saran .....	50
DAFTAR REFERENSI .....	51
LAMPIRAN .....	53



## DAFTAR TABEL

Tabel 2.1 Kajian Pustaka .....	6
Tabel 4.1 Aturan Tekanan Darah .....	19
Tabel 4.2 Penentuan Tekanan Darah .....	20
Tabel 4.3 Nilai Tekanan Darah .....	20
Tabel 4.4 Nilai Diet .....	21
Tabel 4.5 Data Latih .....	22
Tabel 4.6 Data Uji .....	23
Tabel 4.7 Data Latih Numerik.....	24
Tabel 4.8 Data Uji Numerik .....	25
Tabel 4.9 Jarak Antar Data Latih .....	26
Tabel 4.10 Hasil Pengurutan .....	27
Tabel 4.11 Pengambilan Data Latih Sebanyak Nilai K.....	28
Tabel 4.12 Data Latih Kategoris .....	28
Tabel 4.13 Data Uji Kategoris.....	29
Tabel 4.14 Probabilitas Likelihood .....	29
Tabel 4.15 Pengujian Data Kelas Seimbang dan Tidak Seimbang.....	32
Tabel 5.1 Spesifikasi Perangkat Keras .....	33
Tabel 5.2 Spesifikasi Perangkat Lunak .....	33
Tabel 6.1 Hasil Pengujian Sebaran Data Kelas Seimbang $K=1-10$ .....	45
Tabel 6.2 Hasil Pengujian Sebaran Data Kelas Seimbang $K=11-20$ .....	45
Tabel 6.3 Hasil Pengujian Sebaran Data Kelas Seimbang $K=21-30$ .....	46
Tabel 6.4 Hasil Pengujian Sebaran Data Kelas Tidak Seimbang $K=1-10$ .....	46
Tabel 6.5 Hasil Pengujian Sebaran Data Kelas Tidak Seimbang $K=11-20$ .....	47
Tabel 6.6 Hasil Pengujian Sebaran Data Kelas Tidak Seimbang $K=21-30$ .....	47

## DAFTAR GAMBAR

Gambar 3. 1 Diagram Blok Metodologi Penelitian .....	12
Gambar 4. 1 Pohon Perancangan .....	15
Gambar 4. 2 Tahapan Proses Diagnosis Tingkat Risiko Penyakit Stroke .....	16
Gambar 4. 3 Diagram Alir Proses <i>K-Nearest Neighbor</i> .....	17
Gambar 4. 4 Proses Hitung <i>Euclidean</i> .....	18
Gambar 4. 5 Diagram Alir Proses <i>Naïve Bayes</i> .....	19
Gambar 4.6 Perancangan Antarmuka Tampilan Data .....	31
Gambar 4.7 Perancangan Antarmuka Klasifikasi .....	31
Gambar 4.8 Perancangan Antarmuka Pengujian.....	32
Gambar 5.1 Implementasi Antarmuka Tampilan Data .....	43
Gambar 5.2 Implementasi Antarmuka Klasifikasi .....	44
Gambar 5.3 Antarmuka Pengujian .....	44
Gambar 6.1 Grafik Hasil Pengujian Pengaruh Sebaran Data Seimbang .....	48
Gambar 6.2 Grafik Hasil Pengujian Pengaruh Sebaran Data Tidak Seimbang.....	48

## DAFTAR KODE PROGRAM

Kode Program 5.1 Implementasi Algoritme Perhitungan Jarak.....	34
Kode Program 5.2 Implementasi Algoritme Pengurutan Jarak .....	35
Kode Program 5.3 Implementasi Algoritme Prior .....	36
Kode Program 5.4 Implementasi Algoritme Likelihood .....	39
Kode Program 5.5 Implementasi Algoritme Posterior .....	42



## DAFTAR LAMPIRAN

LAMPIRAN ..... 53

