

LEMBAR PERSETUJUAN
IMPLEMENTASI METODE KLUSTERING UNTUK KLASIFIKASI
KANKER PAYUDARA MENGGUNAKAN ALGORITMA NAÏVE BAYES
DAN K-MEANS

SKRIPSI



Disusunoleh:
ANGGREYNI
NIM. 105090606111003

Skripsi ini telah disetujui oleh dosen pembimbing
Pada tanggal 23 April 2015:

Dosen Pembimbing I,

Dosen Pembimbing II,

Dian Eka Ratnawati, S.Si., M.Kom
NIP.197306192002122001

Edy Santoso, S.Si., M.Kom
NIP.197404142003121004

LEMBAR PENGESAHAN
IMPLEMENTASI METODE KLUSTERING UNTUK KLASIFIKASI
KANKER PAYUDARA MENGGUNAKAN ALGORITMA NAÏVE BAYES
DAN K-MEANS

SKRIPSI

Disusun oleh :
ANGGREYNI
NIM. 105090606111003

Skripsi ini telah diuji dan dinyatakan lulus pada tanggal 22 Mei 2015

Dosen Penguji I

Dosen Penguji II

Budi Darma Setiawan, S.Kom., M.Cs.
NIK.

Lailil Muflikhah, S.Kom., M.Sc.
NIK.

Dosen Penguji III

Indriati, S.T., M.Kom.
NIK. 831013 06 1 2 0035

Mengetahui
Ketua Program Studi Teknik Informatika/Illmu komputer

Drs. Mardji, M.T.
NIP. 196708011992031001

KATA PENGANTAR

Puji syukur kepada Allah Yang Maha Esa yang telah melimpahkan anugrah-Nya sehingga Skripsi ini dapat diselesaikan.

Penyusunan Skripsi ini merupakan salah satu syarat kelulusan untuk memperoleh gelar sarjana di FILKOM Universitas Brawijaya. Skripsi ini bertujuan agar mahasiswa dapat mempelajari dan menambah wawasan pengetahuannya tentang hal-hal baru (khususnya dalam bidang teknologi informasi) yang tidak didapatkan dalam kuliah.

Penyelesaian tugas akhir ini tidak lepas dari bantuan dan dorongan berbagai pihak, oleh karena itu saya ingin menyampaikan terima kasih yang setulus-tulusnya kepada :

1. Ibu Dian Eka Ratnawati, S.Si., M.Kom selaku dosen pembimbing I dan bapak Edy Santoso, S.Si., M.Kom selaku dosen pembimbing II atas kesabaran dan bimbingan serta masukan yang diberikan.
2. Drs. Marji, M.T. selaku Ketua Program Studi Teknik Informatika/Ilmu Komputer, Program Teknologi Informasi dan Ilmu Komputer.
3. Segenap bapak dan ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada penulis.
4. Bapak Rinadi dan mamak Emy, Kedua orang tuaku serta keluarga besar yang selalu mendorong, mengingatkan dan memberi semangat.
5. Rekan-rekan penulis, yaitu Eva F. Bisono, S.Kom, Heny Herawati, S.Kom, Jesicha Dwi Ayu Mayasari, S.Kom, dan Alfa Ridhoana, S.Kom yang telah memberikan dukungannya ilmu selama mengerjakan skripsi ini.
6. Teman-teman seangkatan Ilmu Komputer 2010 yang selalu memberikan dukungan.
7. Teman-teman seperantauan, yaitu Nakwin, Onk, Enal, MbK Naponk, Ulan dan Om Ipen yang turut membantu dalam proses penyusunan skripsi ini.

8. Pihak lain yang telah membantu terselesaikannya skripsi ini yang tidak bisa penulis sebutkan satu-persatu.

Semoga segala dorongan, bimbingan, bantuan dan dukungan yang telah diberikan mendapat balasan dari Allah SAW.

Saya menyadari bahwa skripsi ini jauh dari sempurna dan untuk itu tegur sapa, kritik dan saran yang sifatnya konstruktif dari pembaca sangat saya harapkan.

Akhir kata, saya berharap agar skripsi ini dapat bermanfaat bagi yang menggunakannya dan membacanya.

Malang, 2015

ANGGREYNI



LEMBAR PERNYATAAN ORISINALITAS SKRIPSI

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 dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan 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, 23 April 2015

Mahasiswa,

Anggreyni

NIM. 105090606111003

IMPLEMENTASI METODE KLUSTERING UNTUK KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-MEANS

Anggreyni. 2014. Implementasi Metode Klustering Untuk Klasifikasi Kanker Payudara Menggunakan Algoritma Naïve Bayes Dan K-Means. Skripsi Program Studi Teknik Informatika/Illmu Komputer, Fakultas Ilmu Komputer Universitas Brawijaya. Pembimbing: Dian Eka Ratnawati, S.Si., M.Kom. dan Edy Santoso, S.Si., M.Kom.

ABSTRAK

Perkembangan teknologi yang makin pesat telah merambah ke semua bidang aspek kehidupan, tidak terkecuali dalam dunia kesehatan. Dalam dunia kesehatan, sebuah penyakit yaitu kanker payudara mengalami peningkatan penderita di Amerika selama 30 tahun terakhir ini. Awalnya 1 dari 30 perempuan sekarang menjadi 1 dari 8 perempuan. Data kanker payudara hanya akan menjadi sekumpulan data yang tidak berguna jika tidak dilakukan penggalan data terhadapnya. Yang menjadi pertanyaan untuk data-data yang semakin menggunung jumlahnya dari waktu ke waktu itu adalah mau diapakan data-data tersebut. Apakah hanya untuk laporan akhir tahun, kemudian dibuang?. Tentu sayang sekali jika data-data tersebut tidak dimanfaatkan untuk keperluan perusahaan atau instansi itu sendiri. Banyak informasi terpendam yang dapat diambil dari sekumpulan data tersebut sehingga dapat memberikan suatu pengetahuan untuk penentuan kebijakan. Menggali data dapat dilakukan dengan cara mengklasifikasikan data penderita kanker payudara menjadi beberapa kelompok. Dari masalah tersebut pada penelitian ini digunakan pengklusteran K-Means dengan Naïve Bayes. Pada K-Means dilakukan pengelompokan data ke dalam 3 klaster (jinak, mungkin, ganas). Selanjutnya hasil 'mungkin' akan dicari peluangnya dengan menggunakan Naïve Bayes. Penggunaan K-Means dan Naïve Bayes pada dataset kanker payudara memberikan akurasi tertinggi sebesar 97,72% dan akurasi terendah 96,02%. Penggunaan metode ini tepat untuk penentuan kategori kelas data kanker payudara. Terbukti dari hasil perhitungan akurasi pengujian data uji setiap skenario melebihi akurasi standar untuk kelayakan sistem.

Kata Kunci: Kanker payudara, metode klustering, algoritma naïve bayes, algoritma k-means.

IMPLEMENTATION OF THE CLUSTERING METHOD FOR CLASSIFYING BREAST CANCER USING NAÏVE BAYES AND K- MEANS ALGORITHM

Anggreyni. 2014. Implementation Of The Clustering Method For Classifying Breast Cancer Using Naïve Bayes And K-Means Algorithm. Skripsi Program Studi Teknik Informatika/Illmu Komputer, Fakultas Ilmu Komputer Universitas Brawijaya. Pembimbing: Dian Eka Ratnawati, S.Si., M.Kom. dan Edy Santoso, S.Si., M.Kom.

ABSTRACT

Rapid technological development has penetrated into all areas of aspects of life, not least in the world of health. In the medical world, a disease that breast cancer patients in the US has increased over the last 30 years. Originally one of 30 women now become one of 8 women. Breast cancer data will only be a set of data that is useless if it does not do data mining against it. The question for the data that is mounting the amount of time that is what to do with these data. Is it just for the end of year report, then discarded ?. Of course a pity if such data are not used for the purposes of the company or institution itself. Many hidden information that can be retrieved from the data set so as to provide a knowledge for policy. Digging data can be done by classifying the data of breast cancer patients into several groups. Of the problem in this study used the k-means clustering with naïve bayes. In the k-means carried grouping data into 3 clusters (benign, perhaps, malignant). Furthermore, the results 'perhaps' be searched using naïve bayes chances. The use of k-means and naïve bayes on breast cancer dataset provides the highest accuracy of 97.72% and 96.02% lowest accuracy. The use of this method is appropriate for determining breast cancer data class category. Evident from the calculation accuracy of the test data test every scenario exceeds the standards for eligibility system accuracy.

Keywords: Breast cancer, clustering method, naïve bayes algorithm, k-means algorithm.

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
KATA PENGANTAR.....	iii
LEMBAR PERNYATAAN	v
ABSTRAK	vi
ABSTRACT	viii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xii
DAFTAR SOURCE CODE.....	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Kanker	6
2.2.1 Kanker Payudara	7
2.2 Data Mining	10
2.2.1 Pengertian Data Maining	10
2.2.2 Proses Data Mining	11
2.2.3 Fungsionalitas Data Maining	11
2.3 Klasifikasi	13
2.4 <i>Clustering</i>	13
2.5 K-Means.....	14
2.5.1 Pengertian K-Means.....	14

2.5.2	Algoritma K-Means	14
2.5	Naïve Bayes	16
2.6.1	Algoritma Naïve Bayes	16
2.6.2	Metode Klasifikasi Naïve Bayes	17
2.7	Gabungan Naïve Bayes dan K-Means	20
BAB III METODE PENELITIAN DAN PERANCANGAN		24
3.1	Studi Literatur	25
3.2	Data Penelitian	26
3.3	Analisa dan Perancangan Sistem	27
3.3.1	Deskripsi Umum Sistem	27
3.3.2	Perancangan Sistem	28
3.4	Perancangan Antarmuka	41
3.5	Perancangan Pengujian	44
BAB IV IMPLEMENTASI		46
4.1	Lingkungan Implementasi	46
4.1.1	Lingkungan Perangkat Keras	46
4.1.2	Lingkungan Perangkat Lunak	46
4.2	Implementasi Program	47
4.2.1	Proses Input Data Latih	47
4.2.2	Proses Membuat <i>Centroid</i> Baru Secara Acak	47
4.2.3	Proses Menghitung Jarak Data Ke <i>Centroid</i>	48
4.2.4	Proses Meng-update <i>Centroid</i>	49
4.2.5	Proses Mendapatkan Data Latih Untuk Naïve Bayes	49
4.2.6	Proses Menghitung Rata-rata dan Varian	51
4.2.7	Proses Perhitungan Menggunakan Distribusi Gauss	52
4.2.8	Proses Naïve Bayes	53
4.2.9	Proses Pelatihan KMNB(K-Means dan Naïve Bayes)	56
4.2.10	Proses Pengujian KMNB(K-Means dan Naïve Bayes)	57
4.3	Implementasi Antarmuka	58
4.4.1	Antarmuka Utama	58
4.4.2	Antarmuka Aplikasi	58

BAB V PENGUJIAN DAN ANALISIS.....	64
5.1 Pengujian Sistem.....	65
5.1.1 Penentuan Jumlah Data Latih dan Data Uji.....	65
5.1.2 Perbandingan Hasil Akurasi Antara Metode Gabungan K-Means dan Naïve Bayes (KMNB) Dengan Metode K-Means Konvensional.....	65
5.1.3 Pengaruh Jumlah Data Latih Terhadap Data Uji	69
5.2 Analisa Hasil.....	70
5.2.1 Analisa Hasil Pengaruh Jumlah Data Latih Terhadap Data Uji.	70
5.2.2 Analisa Perbandingan Hasil Akurasi Metode Gabungan K-Means dan Naïve Bayes Dengan Metode K-Means Konvensional.....	71
BAB VI PENUTUP	73
6.1 Kesimpulan	73
6.2 Saran	74
DAFTAR PUSTAKA.....	DP-1
LAMPIRAN.....	L-1

DAFTAR GAMBAR

Gambar 2. 1 Aliran informasi dalam <i>data mining</i>	11
Gambar 2. 2 Arsitektur sistem	20
Gambar 2. 3 Proses K-Means	22
Gambar 2. 4 Proses Naïve Bayes	23
Gambar 3. 1 Tahapan penelitian	24
Gambar 3. 2 Tahapan pelatihan dan pengujian pada sistem	28
Gambar 3. 3 Alur sistem pengelompokkan K-Means.....	30
Gambar 3. 4 Alur sistem klasifikasi Naïve Bayes.....	32
Gambar 3. 5 Tampilan awal sistem.....	41
Gambar 3. 6 Tampilan aplikasi K-Means dan Naïve Bayes	43
Gambar 4. 1 Halaman utama.....	59
Gambar 4. 2 Antarmuka untuk memasukkan data latih (KMNB)	60
Gambar 4. 3 Antarmuka pelatihan algoritma K-Means dan Naïve Bayes	61
Gambar 4. 4 Antarmuka untuk memasukkan data uji (KMNB)	62
Gambar 4. 5 Antarmuka pengujian algoritma K-Means dan Naïve Bayes.....	63
Gambar 5. 1 Pengaruh jumlah data latih terhadap data uji	70
Gambar 5. 2 Perbandingan akurasi K-Means dan Naïve Bayes dengan K-Means	71

DAFTAR TABEL

Tabel 3. 1 Atribut dataset kanker payudara	26
Tabel 3. 2 Contoh data latih kanker payudara.....	33
Tabel 3. 3 <i>Centroid</i> awal.....	34
Tabel 3. 4 Hasil perhitungan jarak <i>eucledian</i> iterasi ke-1.....	35
Tabel 3. 5 Klaster baru iterasi ke-2	36
Tabel 3. 6 Hasil perhitungan jarak <i>eucledian</i> iterasi ke-2.....	36
Tabel 3. 7 Klaster baru iterasi ke-3	36
Tabel 3. 8 Hasil perhitungan jarak <i>eucledian</i> iterasi ke-3.....	37
Tabel 3. 9 Data C2	37
Tabel 3. 10 Rata-rata Naïve Bayes.....	38
Tabel 3. 11 Varian Naïve Bayes	38
Tabel 3. 12 Perhitungan distribusi gauss C1.....	39
Tabel 3. 13 Perhitungan distribusi gauss C2.....	39
Tabel 3. 14 Mendapat nilai maksimum.....	40
Tabel 3. 15 Perancangan uji coba sistem	45
Tabel 5. 1 Jumlah data latih dan data uji.....	65
Tabel 5. 2 Hasil akurasi metode K-Means dengan data latih 60%	66
Tabel 5. 3 Hasil akurasi metode K-Means dengan data latih 70%	66
Tabel 5. 4 Hasil akurasi metode K-Means dengan data latih 80%	67
Tabel 5. 5 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 60%	67
Tabel 5. 6 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 70%	68
Tabel 5. 7 Hasil akurasi metode K-Means dan Naïve Bayes dengan data latih 60%	68
Tabel 5. 8 Hasil percobaan pengaruh jumlah data latih dengan data uji.....	69

DAFTAR SOURCE CODE

Source code 4. 1 Proses input data latih.....	47
Source code 4. 2 Proses membuat <i>centroid</i> baru secara acak.....	48
Source code 4. 3 Proses menghitung jarak data ke <i>centroid</i>	48
Source code 4. 4 Proses mengupdate <i>centroid</i>	49
Source code 4. 5 Proses mendapatkan data latih untuk Naïve Bayes	49
Source code 4. 6 Proses mengitung rata-rata dan varian Naïve Bayes	51
Source code 4. 7 Proses perhitungan menggunakan distribusi gauss	52
Source code 4. 8 Proses Naïve Bayes	53
Source code 4. 9 Proses Pelatihan KMNB.....	56
Source code 4. 10 Proses Pengujian KMNB	57



DAFTAR LAMPIRAN

Lampiran 1 Informasi Sumber Dataset Kanker Payudara	L-1
Lampiran 2 Data Hasil Prediksi KMNB, data uji 20% dengan data latih 80%. ..	L-2
Lampiran 3 Data Hasil Prediksi K-Means, data uji 20% dengan data latih 80%.L-4	
Lampiran 4 Biodata Peneliti	L-6

