



**KLASIFIKASI PENYAKIT KANKER PAYUDARA PADA
ALGORITMA *NAÏVE BAYES* DAN *SUPPORT VECTOR
MACHINE* MENGGUNAKAN *INFORMATION GAIN RATIO*
DAN *PARTICLE SWARM OPTIMIZATION***

Skripsi

diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer
Program Studi Teknik Informatika

Oleh

Farhan Aidil Januar
4611417066

**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG
SEMARANG, 2023**

LEMBAR PERSETUJUAN

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi pada :

Hari : Selasa

Tanggal : 18 Juli 2023

Semarang, 18 Juli 2023

Pembimbing



Aji Purwinarko, S.Si., M.Cs.

PENGESAHAN PENGUJI



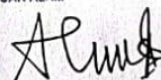



Skripsi berjudul "Klasifikasi Penyakit Kanker Payudara Pada Algoritma *Naive Bayes* Dan *Support Vector Machine* Menggunakan *Information Gain Ratio* Dan *Particle Swarm Optimization*" yang disusun oleh

Nama : Farhan Aidil Januar

Nim : 4611417066

Prodi : Teknik Informatika

Telah dipertahankan dalam ujian skripsi pada hari Jumat, tanggal 1 September 2023.

Ketua Penguji Prof. Dr. Edy Cahyono, M.Si. NIP. 196412051990021001	 
Sekretaris Dr. Alamsyah, S. Si., M. Kom. NIP. 197405172006041001	
Penguji 1 Dr. Alamsyah, S. Si., M. Kom. NIP. 197405172006041001	
Penguji 2 Devi Ajeng Efrianda S.Kom., M.Kom. NIP. 199304152019032012	
Penguji 3/Pembimbing Aji Purwinarko, S.Si., M.Cs. NIP. 198509102015041001	

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila dikemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Semarang, 18 Juli 2023



Farhan Aidil Januar

4611417066

MOTTO

“Maka sesungguhnya bersama kesulitan itu ada kemudahan. Sesungguhnya bersama kesulitan itu ada kemudahan.” (QS. Al Insyirah:5-6)

“Satu-satunya cara untuk melakukan pekerjaan yang hebat adalah dengan mencintai apa yang kamu lakukan.” (Steve Jobs)

“Kunci keberhasilan yang sebenarnya adalah konsistensi.” (B.J. Habibie)

PERSEMBAHAN

Skripsi ini penulis persembahkan kepada kedua orang tua tercinta Ayah Ir. Robi Ampera dan Ibu Eni Murwanti atas cinta, kasih sayang dukungan serta doanya yang selalu menyertai.

Istri Tati Umi Habibah, S.Pd., dan anak Muhammad Arka Haffaz Farhumi yang selalu mendukung dan memberi semangat, serta Kakak-adik tersayang Fajar Hadifadhulloh, S.SE., dan Reni Amalia Nabilah.

Rekan-rekan seperjuangan UNNES, serta almamater kebanggaan saya Universitas Negeri Semarang.

ABSTRAK

Januar, Farhan Aidil. 2023. Klasifikasi Penyakit Kanker Payudara Pada Algoritma *NAÏVE BAYES* Dan *SUPPORT VECTOR MACHINE* Menggunakan *INFORMATION GAIN RATIO* Dan *PARTICLE SWARM OPTIMIZATION*. Skripsi, Teknik Informatika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Aji Purwinarko, S.Si., M.Cs

Kata Kunci: Data Mining, Information Gain Ratio, Naïve Bayes, Support Vector Machine, Particle Swarm Optimization, Klasifikasi, Seleksi Fitur.

Kanker payudara merupakan jenis kanker yang menempati urutan kedua sebagai penyakit yang paling umum ditemui. Kanker secara umum dibagi dua yaitu jinak dan ganas, begitupun kanker payudara. Pada status ganas, kanker dapat berakibat buruk bagi penderitanya bila terlambat diketahui. Oleh sebab itu, deteksi dini pada penyakit kanker sangatlah penting agar penderitanya dapat menerima penanganan yang tepat. Penelitian ini dilakukan dalam rangka untuk melakukan klasifikasi jenis kanker yang mempengaruhi menggunakan teknik data mining. Klasifikasi kanker payudara dilakukan menggunakan metode *data mining* algoritma *naïve bayes* dan *support vector machine* menggunakan seleksi fitur *information gain ratio*. Pada penelitian ini dilakukan dengan menggunakan dataset yang diambil dari www.kaggle.com. 2. Dari hasil penelitian klasifikasi diagnosis penyakit kanker payudara menggunakan Algoritma *naïve bayes* dan SVM dengan menerapkan seleksi fitur IGR yang dilanjutkan dengan PSO didapatkan akurasi terbaik pada Naïve Bayes sebesar 98,24% dan SVM sebesar 96,49%. Dengan demikian, dapat disimpulkan bahwa penerapan algoritma seleksi fitur IGR dan PSO dapat meningkatkan akurasi Algoritma Naïve Bayes yakni sebesar 49,12% dan algoritma SVM yakni sebesar 19,5% pada diagnosis penyakit kanker payudara.

PRAKATA

Puji syukur penulis panjatkan kehadiran Allah SWT karena berkat limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Penyakit Kanker Payudara pada Algoritma *Naïve Bayes* dan *Support Vector Machine* menggunakan *Information Gain Ratio* dan *Particle Swarm Optimization*”.

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan dengan baik tanpa adanya partisipasi dan bantuan dari berbagai pihak. Oleh sebab itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. S Martono, M.Si., Rektor Universitas Negeri Semarang yang telah memberikan kesempatan penulis untuk menyelesaikan masa studi di UNNES;
2. Prof. Edy Cahyono, M.Si., Dekan FMIPA Universitas Negeri Semarang yang telah memberikan izin untuk melaksanakan penelitian;
3. Dr. Alamsyah, S.Si., M.Kom., Koordinator Program Studi Teknik Informatika yang telah memberikan kelancaran dalam penyusunan skripsi;
4. Anggyi Trisnawan Putra, S.Si., M.Si., Dosen Wali yang telah memberikan bimbingan serta motivasi selama masa studi di Prodi Teknik Informatika Unnes;
5. Aji Purwinarko, S.Si., M.Cs., Dosen Pembimbing yang penuh kesabaran memberikan bimbingan, arahan, serta saran kepada penulis selama penyusunan skripsi;
6. Seluruh dosen Prodi Teknik Informatika yang telah memberikan bekal ilmu kepada penulis selama menempuh masa studi di Unnes;
7. Orang tua saya yaitu Bapak Ir. Roby Ampera dan ibu Eni Murwanti, istri serta anak tercinta, kakak-adik serta semua keluarga yang selalu mendukung dan memberi semangat kepada penulis dalam menyelesaikan skripsi;
8. Para sahabat pecel boys (mas ireng, ical, sad boy, jomblo akut, dan si bucin)
9. Fajri, Mahrus, Qois, Alm. Avesina, Surya, Mamed, sanyoto ricky, surya dan zulfan sebagai teman satu kost;

10. Semua pihak yang telah membantu dalam penelitian ini yang tidak dapat penulis sebutkan satu persatu

Semoga penulisan skripsi ini dapat memberikan manfaat dan keberkahan bagi setiap orang yang membacanya

Semarang, 18 Juli 2023

Penulis

Farhan Aidil Januar

4611417066

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN PENGUJI.....	iii
PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRAK	vi
PRAKATA.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian	6
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	7
1.6 Keaslian Penelitian.....	7
BAB 2 TINJAUAN PUSTAKA.....	10
2.1 Tinjauan Pustaka.....	10
2.2 Landasan Teori.....	11
2.2.1 <i>Data Mining</i>	11
2.2.1.1 Pengertian <i>Data Mining</i>	11
2.2.1.2 Tahapan Proses dalam <i>Data Mining</i>	13
2.2.2 Klasifikasi	15

2.2.3	<i>Naïve Bayes</i>	16
2.2.4	<i>Support Vector Machine (SVM)</i>	17
2.2.5	<i>Feature Selection</i>	19
2.2.5.1	<i>Information Gain Ratio (IGR)</i>	20
2.2.5.2	<i>Particle Swarm Optimization (PSO)</i>	21
2.2.6	<i>Confusion Matrix</i>	22
2.2.7	<i>Dataset</i>	23
2.2.8	<i>Breast Cancer Wisconsin Dataset</i>	23
BAB 3	METODE PENELITIAN	25
3.1	Studi Literatur	25
3.2	Pengambilan Data	26
3.3	Pengolahan Data	26
3.3.1	Tahapan <i>Preprocessing</i>	28
3.3.2	Tahapan <i>Feature Selection</i>	28
3.3.3	Tahapan Split Data.....	31
3.3.4	Tahap Klasifikasi	31
3.3.4.1	Klasifikasi Algoritma <i>Naïve Bayes</i>	31
3.3.4.2	Klasifikasi Algoritma <i>SVM</i>	33
3.3.5	Tahap Evaluasi dengan <i>Confusion Matrix</i>	35
3.3.6	Tahap Pengujian Model	36
3.4	Perancangan Sistem	36
3.5	Penarikan Kesimpulan	36
BAB 4	HASIL PENELITIAN DAN PEMBAHASAN	38
4.1	Hasil Penelitian	38
4.1.1	Hasil <i>Preprocessing Data</i>	38
4.1.2	Hasil <i>Feature Selection</i>	40
4.1.2.1	Hasil <i>Feature Selection IGR</i>	40

4.1.2.2	Hasil <i>Feature Selection</i> PSO	42
4.1.3	Hasil <i>Data Mining</i>	43
4.1.3.1	Hasil Klasifikasi Algoritma <i>Naïve</i> Bayes tanpa <i>IGR</i> dan <i>PSO</i>	43
4.1.3.2	Hasil Klasifikasi Algoritma SVM tanpa <i>IGR</i> dan <i>PSO</i>	45
4.1.3.3	Hasil Klasifikasi Algoritma <i>Naïve</i> Bayes dengan <i>IGR</i> dan <i>PSO</i>	48
4.1.3.4	Hasil Klasifikasi Algoritma SVM dengan <i>IGR</i> dan <i>PSO</i>	51
4.1.4	Hasil Implementasi Sistem	54
4.2	Pembahasan.....	57
BAB 5	PENUTUP.....	62
5.1	Simpulan	62
5.2	Saran	63
	DAFTAR PUSTAKA	64

DAFTAR TABEL

Tabel 1.1	Hasil Penelitian Sebelumnya.....	9
Tabel 2.1	Dataset <i>Breast Cancer Wisconsin Dataset</i>	24
Tabel 3.1	Tabel <i>Confusion Matrix</i> 2 Kelas	36
Tabel 4.1	<i>Formatting Heart Failure Prediction Dataset</i>	39
Tabel 4.2	Hasil Pembobotan dengan IGR.....	40
Tabel 4.3	Hasil Akurasi Setiap k Proses IGR pada SVM	41
Tabel 4.4	Hasil Akurasi Setiap k Proses IGR pada Naïve Bayes.....	41
Tabel 4.5	Nilai <i>best cost</i> Naïve Bayes dan SVM	42
Tabel 4.6	Hasil Akurasi Algoritma Naïve Bayes tanpa IGR dan PSO.....	44
Tabel 4.7	<i>Confusion matrix</i> Akurasi Tertinggi Algoritma Naïve Bayes tanpa IGR dan PSO.....	44
Tabel 4.8	<i>Confusion matrix</i> Akurasi Terendah Algoritma Naïve Bayes tanpa IGR dan PSO.....	45
Tabel 4.9	Hasil Akurasi Algoritma SVM tanpa IGR dan PSO	46
Tabel 4.10	<i>Confusion matrix</i> Tertinggi Algoritma SVM tanpa IGR dan PSO ...	47
Tabel 4.11	<i>Confusion matrix</i> Terendah Algoritma SVM tanpa IGR dan PSO ...	47
Tabel 4.12	Hasil Akurasi Algoritma Naïve Bayes dengan IGR dan PSO.....	49
Tabel 4.13	<i>Confusion matrix</i> Akurasi Tertinggi Algoritma Naïve Bayes dengan IGR dan PSO	50
Tabel 4.14	<i>Confusion matrix</i> Akurasi Terendah Algoritma Naïve Bayes dengan IGR dan PSO	50
Tabel 4.15	Hasil Akurasi Algoritma SVM tanpa IGR dan PSO	52
Tabel 4.16	<i>Confusion matrix</i> Tertinggi Algoritma SVM dengan IGR dan PSO.	53
Tabel 4.17	<i>Confusion matrix</i> Terendah Algoritma SVM dengan IGR dan PSO	53
Tabel 4.18	Hasil Akurasi Akhir Percobaan pada Algoritma Naïve Bayes dan SVM	59
Tabel 4.19	Hasil Akurasi Akhir Percobaan pada Algoritma Naïve Bayes dan SVM dengan IGR dan PSO.....	59

Tabel 4.20 Hasil Akhir Tertinggi Akurasi Penelitian.....	59
Tabel 4.21 Hasil Penelitian Sebelumnya.....	60

DAFTAR GAMBAR

Gambar 2.1	Proses <i>Data Mining</i>	14
Gambar 3.1	Tahapan Penelitian	25
Gambar 3.2	<i>Flowchart</i> Metode yang diusulkan.....	27
Gambar 3.3	<i>Flowchart</i> Algoritma <i>IGR</i>	29
Gambar 3.4	<i>Flowchart</i> Algoritma <i>PSO</i>	30
Gambar 3.5	<i>Flowchart</i> Algoritma <i>Naïve Bayes</i>	33
Gambar 3.6	<i>Flowchart</i> Algoritma <i>SVM</i>	35
Gambar 4.1	<i>Breast Cancer Wisconsin Dataset</i> dengan format <i>.arrrf</i>	38
Gambar 4.2	<i>Breast Cancer Wisconsin Dataset</i> dengan format <i>CSV</i>	39
Gambar 4.3	Potongan Program Algoritma <i>Naïve Bayes</i> tanpa <i>IGR</i> dan <i>PSO</i>	43
Gambar 4.4	Potongan Program Algoritma <i>SVM</i> tanpa <i>IGR</i> dan <i>PSO</i>	46
Gambar 4.5	Potongan Program Algoritma <i>Naïve Bayes</i> dengan <i>IGR</i> dan <i>PSO</i>	48
Gambar 4.6	Potongan Program Algoritma <i>SVM</i> dengan <i>IGR</i> dan <i>PSO</i>	52
Gambar 4.7	Tampilan Menu pada <i>Dashboard</i>	54
Gambar 4.8	Tampilan <i>Input Dataset</i> pada Menu <i>Dataset</i>	55
Gambar 4.9	Hasil <i>Dataset</i> yang diunggah pada Menu <i>Dataset</i>	55
Gambar 4.10	Hasil Transformasi Data pada Menu <i>Dataset</i>	56
Gambar 4.11	Tampilan Menu Akurasi.....	56
Gambar 4.12	Tampilan Menu <i>About</i>	56
Gambar 4.12	Perbandingan Akurasi dengan Penelitian Sebelumnya	61

DAFTAR LAMPIRAN

Lampiran 1	Hasil Pengumpulan Data	71
Lampiran 2	Hasil Information Gain Ratio	72
Lampiran 3	Hasil Akurasi Setiap k Proses <i>IGR</i> pada <i>SVM</i>	73
Lampiran 4	Hasil Akurasi Setiap k Proses <i>IGR</i> pada <i>Naïve Bayes</i>	74
Lampiran 5	Hasil Akurasi Algoritma <i>Naïve Bayes</i> Tanpa <i>IGR</i> dan <i>PSO</i>	75
Lampiran 6	Hasil Akurasi Algoritma <i>SVM</i> Tanpa <i>IGR</i> dan <i>PSO</i>	76
Lampiran 7	Hasil Akurasi Algoritma <i>Naïve Bayes</i> dengan <i>IGR</i> dan <i>PSO</i>	77
Lampiran 8	Hasil Akurasi Algoritma <i>SVM</i> dengan <i>IGR</i> dan <i>PSO</i>	78

