

**“PENGARUH SELEKSI FITUR CHI SQUARE
TERHADAP KLASIFIKASI
NAÏVE BAYES PADA DATA KANKER PAYUDARA”**

Emmilia Norhafifah
(170411200135)

Pembimbing I : Mula'ab S.Si.,M.Kom.

Pembimbing II : Imamah S.Kom.,M.Kom

ABSTRAK

Kanker Payudara adalah tumor ganas yang menyerang jaringan payudara yang berasal dari kelenjar, saluran kelenjar dan jaringan penunjang payudara. Kanker payudara terjadi karena adanya kerusakan gen yang mengatur pertumbuhan dan diferensiasi sehingga sel ini tumbuh dan berkembang biak tanpa dapat dikendalikan. Kanker payudara menjadi pembunuh wanita terbanyak di dunia. Pencatatan terhadap penyakit kanker banyak dilakukan guna mengantisipasi dan menganalisa pasien sejak dini agar dapat dilakukan pencegahan. Salah satu algoritma klasifikasi yang dapat menangani data dalam jumlah besar adalah *Naïve Bayes*. data dalam jumlah besar yang memiliki banyak fitur menimbulkan beberapa masalah salah satunya semakin lama proses klasifikasi. Tujuan dari penelitian ini untuk mengetahui pengaruh penambahan seleksi fitur menggunakan *chi square* terhadap klasifikasi *Naïve Bayes* pada data kanker payudara. Hasil yang didapat dari penelitian ini adalah nilai akurasi sebesar 0.962, presisi 0.955, *recall* sebesar 0.965 dan *f-measure* sebesar 0.959. Akurasi tertinggi didapat pada klasifikasi dengan 8 fitur.

Kata kunci : kanker payudara, klasifikasi teks, *chi square*, Naive Bayes, seleksi fitur.

KATA PENGANTAR

Alhamdulillah puji syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “**Pengaruh Seleksi Fitur *Chi Square* Terhadap Klasifikasi *Naïve Bayes* Pada Data Kanker Payudara**”. Dengan terselesaikannya laporan skripsi ini, penulis tak lupa menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah melimpahkan rahmat dan keberkahan bagi penulis sehingga penulis dapat merasakan nikmat yang diberikan sampai saat ini. Dan Nabi Muhammad SAW yang telah menuntun umat manusia dari kebodohan menuju cahaya Islam.
2. Kedua orang tua penulis, Ibu Kamsiyah dan Bapak Ach Subaidi serta keluarga penulis yang selalu memberikan doa, semangat, kasih sayang dan motivasi baik moral maupun materi. Berkat do'a mereka penulis bisa menyelesaikan penelitian ini. Mereka pula yang senantiasa menyumbang semangat saat penulis rapuh dalam segala hal.
3. Bapak Mula'ab S.Si.,M.Kom. dan ibu Imamah S.Kom.,M.Kom. selaku dosen pembimbing yang selalu memberikah arahan dan bimbingan selama penelitian skripsi ini.
4. Segenap staf dan dosen di Teknik Informatika yang telah membimbing penulis selama masa perkuliahan.
5. Rekan-rekan Alih jenjang teknik informatika 2017 yang senantiasa membantu dalam keadaan suka dan duka bersama penulis. serta Semua teman-teman penulis yang tidak dapat di sebutkan satu persatu yang selalu memberi dukungan dan semangat.

Penulis menyadari bahwa pengerjaan skripsi ini tidak luput dari berbagai kesalahan. Oleh karena itu, penulis menyatakan permohonan maaf atas segala kekurangan yang ada. Dengan kerendahan hati, berbagai saran dan kritik yang membangun sangat penulis harapkan dari rekanrekan pembaca.

Bangkalan, 18 Juli 2019.

Penulis

DAFTAR ISI

ABSTRAK	v
KATA PENGANTAR	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Permasalahan	3
1.2.1 Permasalahan	3
1.2.2 Usulan Solusi	3
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan dan Manfaat	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	4
1.4 Batasan Masalah	4
1.5 Sistematika Penulisan	4
BAB II STUDI PUSTAKA	6
2.1 Landasan Teori	6
2.1.1 Kanker Payudara	6
2.1.2 Seleksi <i>Fitur</i>	8
2.1.3 Chi Square	9
2.1.4 Naïve Bayes	9
2.1.5 K-fold Cross Validation	11
2.1.6 Penelitian Sebelumnya	13
BAB III METODE USULAN	17
3.1 Studi Literatur	17
3.2 Analisis	17
3.2.1 Analisis Proses	17
3.2.2 Analisis Output	17
3.2.3 Rancangan (Arsitektur) Sistem	17
3.3 Pengujian dan Evaluasi	18
3.4 Contoh Perhitungan	19
BAB IV HASIL DAN PEMBAHASAN	23
4.1 Spesifikasi Perangkat Lunak	23

4.2 Pembuatan Program	23
4.2.1 Pembuatan GUI	23
4.2.2 Memuat Dataset.....	25
4.2.3 Seleksi Fitur.....	25
4.2.4 Model Klasifikasi	26
4.2.5 Evaluasi	26
4.3 Evaluasi dan Hasil Analisa Sistem.....	26
BAB V PENUTUP.....	31
5.1 Kesimpulan.....	31
5.2 Saran	31
REFERENSI	32

DAFTAR TABEL

Tabel 1.1 Perumusan Masalah	3
Tabel 2.1 <i>Fitur Breast Cancer Wiscosin (Original)</i>	7
Tabel 2.2 Model <i>Confusion matrix</i>	12
Tabel 2.3 Peneliti sebelumnya	14
Tabel 2.4 Dataset.....	19
Tabel 2.5 Frekuensi observasi.....	20
Tabel 2.6 Frekuensi harapan	20
Tabel 2.7 Chi hitung.....	21
Tabe 4.1 Spesifikasi perangkat Lunak	23
Tabel 4.2 Hasil Percobaan Skenario Uji Coba 1.....	29
Tabel 4.3 Hasil Percobaan Skenario 2	29
Tabel 4.4 Perbandingan Hasil Percobaan	29

DAFTAR GAMBAR

Gambar 2.1 Skema <i>k-fold cross validation</i>	11
Gambar 3.1 Rancangan Sistem	17
Gambar 3.2 Flowchart Rancangan Evaluasi	18
Gambar 4.1 Rancangan GUI Sistem	23
Gambar 4.2 Tab <i>Loading</i> Dataset	24
Gambar 4.3 Tab Seleksi Fitur	24
Gambar 4. 4 Tab Klasifikasi	25
Gambar 4.5 Uji Coba Memuat Dataset	27
Gambar 4.6 Uji Coba Seleksi Fitur	27
Gambar 4.7 Uji Coba Klasifikasi	28