



PENERAPAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* DAN SELEKSI FITUR *GENETIC ALGORITHM* PADA KLASIFIKASI DATA MICROARRAY MENGGUNAKAN *SUPPORT VECTOR MACHINE*

SKRIPSI

Untuk memenuhi persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh

RAHMAT AMIN BADALI

2011016210009

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
DESEMBER 2024**



**PENERAPAN REDUKSI DIMENSI *PRINCIPAL COMPONENT*
ANALYSIS DAN SELEKSI FITUR *GENETIC ALGORITHM* PADA
KLASIFIKASI DATA MICROARRAY MENGGUNAKAN *SUPPORT*
*VECTOR MACHINE***

SKRIPSI

Untuk memenuhi persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh
RAHMAT AMIN BADALI
2011016210009

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
DESEMBER 2024

SKRIPSI

PENERAPAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS DAN SELEKSI FITUR GENETIC ALGORITHM PADA KLASIFIKASI DATA MICROARRAY MENGGUNAKAN SUPPORT VECTOR MACHINE


Oleh :

RAHMAT AMIN BADALI


NIM. 2011016210009

Telah dipertahankan di depan Dosen Penguji pada tanggal 24 Desember 2024.
Susunan Dosen Penguji :


Pembimbing I


Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003


Dosen Penguji I


Dodon Turianto Nugrahadi, S.Kom., M.Eng.
NIP. 198001122009121002

Pembimbing II


Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002


Dosen Penguji II


Setvo Wahyu Saputro, S.Kom., M.Kom.
NIP. 198808072023211027



Banjarbaru, 31 Desember 2024

Koordinator Program Studi Ilmu Komputer,


Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya jua tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 31 Desember 2024

Yang menyatakan,



Rahmat Amin Badali

NIM. 2011016210009

ABSTRAK

PENERAPAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* DAN SELEKSI FITUR *GENETIC ALGORITHM* PADA KLASIFIKASI DATA MICROARRAY MENGGUNAKAN *SUPPORT VECTOR MACHINE* (Oleh: Rahmat Amin Badali; Pembimbing: Dwi Kartini, S.Kom., M.Kom. dan Muliadi, S.Kom., M.Cs.;2024;127 halaman)

DNA microarray telah berkembang pesat dan menjadi alat penting dalam penelitian bioinformatika. Teknologi ini memungkinkan pemantauan ekspresi gen dalam jumlah besar secara bersamaan, sehingga memberikan manfaat yang signifikan dalam berbagai eksperimen biologi, termasuk salah satunya deteksi kanker. Namun, data microarray memiliki kelemahan berupa dimensi data yang tinggi dan jumlah sampel yang relative sedikit. Hal ini dapat mempengaruhi akurasi pada model klasifikasi dalam mendeteksi ekspresi gen. Pada penelitian ini digunakan Principal Component Analysis (PCA) sebagai metode mereduksi dimensi dan Genetic Algorithm (GA) sebagai metode seleksi fitur serta klasifikasi Support Vector Machine (SVM) untuk mengklasifikasi data microarray. Semua metode ini diterapkan pada tiga dataset microarray yaitu *breast cancer*, *ovarian cancer*, dan leukemia. Berdasarkan hasil pengujian, diketahui bahwa penerapan reduksi dimensi PCA dengan seleksi fitur GA pada klasifikasi SVM mampu mengurangi kompleksitas data dan meningkatkan akurasi model klasifikasi. Hasil klasifikasi pada dataset breast cancer didapatkan akurasi sebesar 73.33% dengan 28 fitur. Pada dataset ovarian cancer didapatkan akurasi sebesar 98.68% dengan 11 fitur. Sedangkan pada dataset leukemia didapatkan akurasi sebesar 95.45% dengan 9 fitur.

Kata kunci: *Microarray, Reduksi Dimensi, Seleksi Fitur, Klasifikasi, Principal Component Analysis, Genetic Algorithm, Support Vector Machine.*

ABSTRACT

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS DIMENSION REDUCTION AND GENETIC ALGORITHM FEATURE SELECTION ON MICROARRAY DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE

(By: Rahmat Amin Badali; Supervisor: Dwi Kartini, S.Kom., M.Kom. ; Muliadi, S.Kom., M.Cs.; 2024; 127 pages)

DNA microarrays have developed rapidly and become an essential tool in bioinformatics research. This technology enables the simultaneous monitoring of gene expression on a large scale, providing significant benefits in various biological experiments, including cancer detection. However, microarray data have limitations in the form of high-dimensional data (a large number of features) and relatively small sample sizes. This can affect the accuracy of classification models in detecting gene expression. In this study, Principal Component Analysis (PCA) was used as a dimensionality reduction method, Genetic Algorithm (GA) as a feature selection method, and Support Vector Machine (SVM) for classifying microarray data. These methods were applied to three microarray datasets: breast cancer, ovarian cancer, and leukemia. The results show that the application of PCA for dimensionality reduction combined with GA for feature selection in SVM classification can reduce data complexity and improve the accuracy of the classification model. For the breast cancer dataset, an accuracy of 73.33% was achieved with 28 features. The ovarian cancer dataset achieved an accuracy of 98.68% with 11 features, while the leukemia dataset achieved an accuracy of 95.45% with 9 features.

Keywords: Microarray, Dimensionality Reduction, Feature Selection, Classification, Principal Component Analysis, Genetic Algorithm, Support Vector Machine.

PRAKATA

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas segala limpahan rahmat, kemudahan, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Reduksi Dimensi Principal Component Analysis dan Seleksi Fitur Genetic Algorithm Pada Klasifikasi Data Microarray Menggunakan Support Vector Machine” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang telah mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Kedua orang tua dan keluarga yang telah memberikan semangat, doa, dan dukungan terhadap penulis untuk menyelesaikan skripsi ini dengan baik.
2. Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen pembimbing utama dan Koordinator Program Studi Ilmu Komputer FMIPA ULM yang telah bersedia meluangkan waktunya untuk memberikan bimbingan, arahan, bantuan serta izin sehingga skripsi ini dapat diselesaikan.
3. Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing pendamping yang juga telah bersedia senantiasa meluangkan waktunya untuk memberikan bimbingan dan arahan selama proses penyelesaian skripsi ini.
4. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Bapak Dodon Turianto Nugrahadi, S.Kom., M.Eng. dan Setyo Wahyu Saputro, S.Kom., M.Kom. selaku tim penguji yang juga membantu menguji, memberi saran dan arahan pada penelitian ini.
6. Teman-teman keluarga Ilmu Komputer Angkatan 2020 yang telah memberikan dukungan dan bantuan kepada penulis untuk bisa menyelesaikan skripsi ini.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang juga telah turut serta membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Oleh karena itu penulis mengharapkan adanya kritik dan saran yang

membangun. Semoga tulisan ini dapat memberikan manfaat bagi ilmu pengetahuan dan bagi para pembaca.

Banjarbaru, 31 Desember 2024

A handwritten signature in black ink, consisting of several loops and a trailing line, positioned above the printed name.

Rahmat Amin Badali

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN	iv
ABSTRAK	v
ABSTRACT.....	vi
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang	16
1.2 Rumusan Masalah	17
1.3 Batasan Masalah.....	18
1.4 Tujuan Penelitian.....	18
1.5 Manfaat Penelitian.....	19
BAB II TINJAUAN PUSTAKA.....	20
2.1 Kajian Terdahulu.....	20
2.2 Keaslian Penelitian.....	24
2.3 Landasan Teori.....	31
2.3.1 Microarray.....	31
2.3.2 Data Mining	32
2.3.3 Normalisasi Data.....	33
2.3.4 Principal Component Analysis.....	34
2.3.5 Genetic Algorithm.....	37
2.3.6 Support Vector Machine	39
2.3.7 Evaluasi Algoritma.....	41
BAB III METODE PENELITIAN.....	43
3.1 Alat Penelitian	43
3.2 Bahan Penelitian.....	43
3.3 Alur Penelitian.....	43
BAB IV HASIL DAN PEMBAHASAN	47

4.1 Hasil	47
4.1.1 Pengumpulan Data	47
4.1.2 Preprocessing Data	49
4.1.3 Pembagian Dataset.....	52
4.1.4 Klasifikasi Support Vector Machine.....	53
4.1.5 Reduksi Dimensi Principal Component Analysis dan Klasifikasi Support Vector Machine.....	55
4.1.6 Seleksi Fitur Genetic Algorithm dan Klasifikasi Support Vector Machine	67
4.1.7 Reduksi Dimensi Principal Component Analysis, Seleksi Fitur Genetic Algorithm dan Klasifikasi Support Vector Machine.	79
4.1.8 Evaluasi.....	93
4.2 Pembahasan.....	100
BAB V PENUTUP.....	111
5.1 Kesimpulan.....	111
5.2 Saran	112
DAFTAR PUSTAKA	113
LAMPIRAN.....	118

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	24
Tabel 2. Rancangan Penelitian.....	30
Tabel 3. Confussion Matrix.....	41
Tabel 4. Dataset Breast Cancer.....	47
Tabel 5. Dataset Ovarian Cancer.....	48
Tabel 6. Dataset Leukemia.....	49
Tabel 7. Label Encoding pada Setiap Dataset.....	50
Tabel 8. Perhitungan Min-Max Normalization Dataset Breast.....	51
Tabel 9. Min-Max Normalization di Setiap Fitur pada Dataset Breast Cancer.....	51
Tabel 10. Min-Max Normalization di Setiap Fitur pada Dataset Ovarian Cancer.....	52
Tabel 11. Min-Max Normalization di Setiap Fitur pada Dataset Leukemia.....	52
Tabel 12. Pembagian Data Training dan Data Testing untuk setiap Dataset.....	53
Tabel 13. Akurasi Klasifikasi SVM.....	54
Tabel 14. Nilai Mean dari Setiap Fitur pada Dataset Breast.....	55
Tabel 15. Perhitungan Centered Data pada Dataset Breast.....	56
Tabel 16. Matriks Kovarian pada Dataset Breast.....	56
Tabel 17. Matriks Kovarian pada Dataset Ovarian.....	57
Tabel 18. Matriks Kovarian pada Dataset Leukemia.....	57
Tabel 19. Eigen Value pada Dataset Breast.....	58
Tabel 20. Eigen Value pada Dataset Ovarian.....	59
Tabel 21. Eigen Value pada Dataset Leukemia.....	59
Tabel 22. Eigen Vector pada Dataset Breast.....	60
Tabel 23. Eigen Vector pada Dataset Ovarian.....	60
Tabel 24. Eigen Vector pada Dataset Leukemia.....	61
Tabel 25. Ratio Proporsi Variance pada Dataset Breast.....	61
Tabel 26. Ratio Proporsi Variance pada Dataset Ovarian.....	62
Tabel 27. Ratio Proporsi Variance pada Dataset Leukemia.....	63
Tabel 28. Jumlah Principal Component yang Dipertahankan Untuk Setiap Variance.....	63
Tabel 29. Reduksi Dimensi PCA pada dataset Breast.....	64
Tabel 30. Reduksi Dimensi PCA pada dataset Ovarian.....	64

Tabel 31. Reduksi Dimensi PCA pada dataset Leukemia.....	65
Tabel 32. Akurasi Klasifikasi SVM Pada Dataset Breast Cancer.....	66
Tabel 33. Akurasi Klasifikasi SVM Pada Dataset Ovarian Cancer	66
Tabel 34. Akurasi Klasifikasi SVM Pada Dataset Leukemia	66
Tabel 35. Paramater Genetic Algorithm	68
Tabel 36. Inisialisasi Populasi Awal Pada Dataset Breast	69
Tabel 37. Nilai Fitness Populasi Awal Pada Dataset Breast.....	70
Tabel 38. Crossover Pada Populasi Awal Dataset Breast	71
Tabel 39. Individu Baru Hasil Crossover.....	71
Tabel 40. Individu Baru Hasil Mutasi.....	72
Tabel 41. Populasi Baru yang Terbentuk.....	73
Tabel 42. Ranking Individu Terbaik pada Generasi Pertama	75
Tabel 43. Akurasi Klasifikasi SVM pada Dataset Breast	78
Tabel 44. Akurasi Klasifikasi SVM pada Dataset Ovarian.....	79
Tabel 45. Akurasi Klasifikasi SVM pada Dataset Leukemia	79
Tabel 46. Reduksi Dimensi PCA pada dataset Breast	80
Tabel 47. Reduksi Dimensi PCA pada dataset Ovarian.....	80
Tabel 48. Reduksi Dimensi PCA pada dataset Leukemia.....	81
Tabel 49. Inisialisasi Populasi Awal Pada Dataset Breast	82
Tabel 50. Nilai Fitness Populasi Awal Pada Dataset Breast.....	83
Tabel 51. Crossover Pada Populasi Awal Dataset Breast.....	84
Tabel 52. Individu Baru Hasil Crossover.....	85
Tabel 53. Individu Baru Hasil Mutasi.....	86
Tabel 54. Populasi Baru yang Terbentuk.....	87
Tabel 55. Ranking Individu Terbaik pada Generasi Pertama	88
Tabel 56. Akurasi Klasifikasi SVM pada Dataset Breast	92
Tabel 57. Akurasi Klasifikasi SVM pada Dataset Ovarian.....	92
Tabel 58. Akurasi Klasifikasi SVM pada Dataset Leukemia	92
Tabel 59. Confussion Matrix Klasifikasi SVM pada Dataset Breast.....	93
Tabel 60. Confussion Matrix Klasifikasi SVM pada Dataset Ovarian	94
Tabel 61. Confussion Matrix Klasifikasi SVM pada Dataset Leukemia	94
Tabel 62. Confussion Matrix Reduksi Dimensi PCA dan Klasifikasi SVM pada	

Dataset Breast.....	95
Tabel 63. Confussion Matrix Reduksi Dimensi PCA dan Klasifikasi SVM pada Dataset Ovarian.....	95
Tabel 64. Confussion Matrix Reduksi Dimensi PCA dan Klasifikasi SVM pada Dataset Leukemia.....	96
Tabel 65. Confussion Matrix Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Breast.....	97
Tabel 66. Confussion Matrix Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Ovarian	97
Tabel 67. Confussion Matrix Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Leukemia	97
Tabel 68. Confussion Matrix Reduksi Dimensi PCA, Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Breast	99
Tabel 69. Confussion Matrix Reduksi Dimensi PCA, Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Ovarian.....	99
Tabel 70. Confussion Matrix Reduksi Dimensi PCA, Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM pada Dataset Leukemia	99
Tabel 71. Akurasi Klasifikasi SVM.....	101
Tabel 72. Akurasi Klasifikasi SVM dengan Reduksi Dimensi PCA pada Setiap Dataset	102
Tabel 73. Akurasi Klasifikasi SVM dengan Seleksi Fitur Genetic Algorithm pada Setiap Dataset.....	104
Tabel 74. Akurasi Klasifikasi SVM dengan Reduksi Dimensi PCA dan Seleksi Fitur Genetic Algorithm pada Setiap Dataset	105
Tabel 75. Peningkatan Akurasi Untuk Setiap Pengujian Dibandingkan Pengujian Klasifikasi SVM.....	108

DAFTAR GAMBAR

Gambar 1. Gene Chip.....	31
Gambar 2. Proses Hibridasi DNA Microarray	32
Gambar 3. Alur Penelitian.....	44
Gambar 4. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Breast	76
Gambar 5. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Ovarian	77
Gambar 6. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Leukemia	77
Gambar 7. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Breast	90
Gambar 8. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Ovarian	90
Gambar 9. Grafik Nilai Fitness Untuk Setiap Generasi pada Dataset Leukemia	91
Gambar 10. Grafik Perbandingan Metrik Evaluasi pada model Klasifikasi SVM	94
Gambar 11. Grafik Perbandingan Metrik Evaluasi pada Model Reduksi Dimensi PCA dan Klasifikasi SVM.....	96
Gambar 12. Grafik Perbandingan Metrik Evaluasi pada Model Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM	98
Gambar 13. Grafik Perbandingan Metrik Evaluasi pada Model Reduksi Dimensi PCA, Seleksi Fitur Genetic Algorithm dan Klasifikasi SVM.....	100
Gambar 14. Perbandingan Nilai Metrik Evaluasi Tertinggi pada Setiap Pengujian Untuk Masing-Masing Dataset	107

DAFTAR LAMPIRAN

Lampiran 1. Source Kode Import Data	119
Lampiran 2. Source Kode Preprocessing Data	119
Lampiran 3. Source Kode Pembagian Data	119
Lampiran 4. Source Kode Reduksi Dimensi Principal Component Analysis.....	120
Lampiran 5. Source Kode Seleksi Fitur Genetic Algorithm.....	120
Lampiran 6. Source Kode Klasifikasi Support Vector Machine.....	126
Lampiran 7. Source Kode Evaluasi Model	126