

PENGARUH BORUTA, RELIEFF, DAN RFECV DALAM KLASIFIKASI KANKER PAYUDARA BERDASARKAN DATA *MICROARRAY*

AHMED SIDDIQ ARRAFI IBN SAKTI



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pengaruh Boruta, ReliefF, dan RFECV dalam Klasifikasi Kanker Payudara Berdasarkan Data *Microarray*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Januari 2025

Ahmed Siddiq Arrafi Ibn Sakti
G6401201117

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

AHMED SIDDIQ ARRAFI IBN SAKTI. Pengaruh Boruta, ReliefF, dan RFECV dalam Klasifikasi Kanker Payudara Berdasarkan Data Microarray. Dibimbing oleh TOTO HARYANTO .

Kanker adalah penyakit mematikan akibat pertumbuhan sel abnormal yang tak terkendali, dan kanker payudara merupakan jenis yang paling umum di Indonesia dengan berbagai *subtype*. Penelitian ini mengklasifikasi *subtype* kanker payudara menggunakan data *microarray* dan *machine learning*, mengingat data *microarray* menghasilkan Data dimensi tinggi yang rentan terhadap *overfitting*. Untuk mengatasi masalah ini, digunakan tiga algoritma *feature selection*: ReliefF, RFECV, dan Boruta. Hasil penelitian menunjukkan bahwa Boruta memberikan performa terbaik pada beberapa model, dengan akurasi tertinggi pada model KNN (0,967) dan SVM (0,987). Relief juga memberikan peningkatan akurasi yang signifikan, terutama pada model SVM dengan akurasi sebesar 0,987, sedangkan RFECV mencatat akurasi terbaik pada model Decision Tree (0,821). Selain performa, Boruta mencatat *runtime* tercepat (603 detik) dengan 901 fitur terseleksi. Hasil ini menjadikan Boruta pilihan unggul untuk klasifikasi *subtype* kanker payudara, sekaligus acuan bagi penelitian lanjutan.

Kata Kunci: *Feature Selection, Kanker, Machine Learning, Microarray*

ABSTRACT

AHMED SIDDIQ ARRAFI IBN SAKTI. The Effect of Boruta, ReliefF, dan RFECV on Classifying Breast Cancer Based on Microarray Data. Supervised by TOTO HARYANTO.

Cancer is a deadly disease caused by the uncontrolled growth of abnormal cells, and breast cancer is the most common type in Indonesia with various subtypes. This study classifies breast cancer subtypes using microarray data and machine learning, considering that microarray data produces high-dimensional data that is prone to overfitting. To address this issue, three feature selection algorithms were used: ReliefF, RFECV, and Boruta. The study results show that Boruta provides the best performance across several models, with the highest accuracy achieved in the KNN (0.967) and SVM (0.987) models. ReliefF also significantly improves accuracy, particularly in the SVM model, with an accuracy of 0.987, while RFECV records the best accuracy in the Decision Tree model (0.821). In addition to performance, Boruta achieves the fastest runtime (603 seconds) with 901 selected features. These findings establish Boruta as the superior choice for breast cancer subtype classification and serve as a reference for future research.

Keywords: *Cancer, Feature Selection, Machine Learning, Microarray*



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PENGARUH BORUTA, RELIEFF, DAN RFECV DALAM KLASIFIKASI KANKER PAYUDARA BERDASARKAN DATA *MICROARRAY*

AHMED SIDDIQ ARRAFI IBN SAKTI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:

1. Dr. Mushthofa, S.Kom., M.Sc.
2. Hari Agung Adrianto, S.Kom., M.Si., Ph.D.



Judul Skripsi : Pengaruh Boruta, ReliefF, dan RFECV dalam Klasifikasi Kanker Payudara Berdasarkan Data Microarray

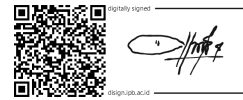
Nama : Ahmed Siddiq Arrafi Ibn Sakti

NIM : G6401201117

Disetujui oleh

Pembimbing 1:

Dr. Toto Haryanto, S.Kom., M.Si.



Diketahui oleh

Ketua Program Sarjana:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom.

198108092008121002



Tanggal Ujian:
30 Desember 2024

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan November 2023 sampai bulan Oktober 2024 ini ialah *machine learning*, dengan judul “Pengaruh Boruta, ReliefF, dan RFECV dalam Klasifikasi Kanker Payudara Berdasarkan Data *Microarray*”.

Penulis menyadari bahwa skripsi ini tidak mungkin selesai tanpa adanya dukungan, bantuan, bimbingan, dan saran dari berbagai pihak selama penyusunan. Pada kesempatan ini, penulis menyampaikan rasa terima kasih kepada:

- a) Ibu serta kedua kakak yang selalu memberikan dukungan tanpa henti setiap kondisi yang saya alami. Terima kasih atas doa, kasih sayang serta pengorbanan yang tiada tara.
- b) Bapak Dr. Toto Haryanto, S.Si., M.Si. selaku dosen pembimbing skripsi yang senantiasa memberikan bimbingan, saran, dan mendorong penulis agar dapat segera menyelesaikan penelitian.
- c) Bapak Dr. Mushthofa, S.Kom., M.Sc. dan Bapak Hari Agung Adrianto, S.Kom., M.Si., Ph. D selaku dosen penguji ujian akhir penelitian yang telah memberikan saran sehingga penelitian ini menjadi lebih baik.
- d) Fakhri Givran yang sudah membantu saya dalam banyak hal selama penelitian ini.
- e) Rekan tim *Valorant* yang sudah menemani penulis saat bermain dengan memberikan motivasi untuk menulis yang sangat keras.
- f) Mahasiswa program studi S1 Ilmu Komputer angkatan 57 yang telah memberikan bantuan dan semangat kepada penulis selama perkuliahan, dan
- g) Valve Corp. dan Riot Games yang memberikan hiburan bagi penulis saat penulis merasa jenuh selama penelitian dan penulisan tugas akhir.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan..

Bogor, Januari 2025

Ahmed Siddiq Arrafi Ibn Sakti



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 Kanker Payudara	4
2.2 Microarray	5
2.3 <i>Feature Selection</i>	6
2.4 Classifier	7
2.5 Boruta	7
2.6 RFECV	8
2.7 ReliefF	9
2.8 SVM	11
2.9 Decision Tree	11
2.10 Random Forest	12
2.11 KNN	13
2.12 Cross-validation	13
III METODE	14
3.1 Data Penelitian	14
3.2 Tahapan Penelitian	14
3.3 Praproses Data	15
3.4 CrossValidation	17
3.5 Classifier	17
3.6 Evaluasi Model	18
IV HASIL DAN PEMBAHASAN	19
4.1 Algoritma Feature Selection	19
4.2 Model Klasifikasi	22
4.3 Perbandingan dengan Penelitian Sebelumnya.	25
4.4 Analisis Kualitatif PAM50	26
V SIMPULAN DAN SARAN	28
5.1 Simpulan	28
5.2 Saran	28
DAFTAR PUSTAKA	29
RIWAYAT HIDUP	35



DAFTAR TABEL

1	Top 10 Fitur ReliefF	19
2	Top 10 Fitur RFECV	20
3	Hasil Seleksi Fitur	21
4	Hasil Perbandingan Penelitian	26
5	Perbandingan Hasil PAM50	27

DAFTAR GAMBAR

1	Ilustrasi cara kerja Boruta (Hasan <i>et al.</i> 2020)	8
2	Ilustrasi alur kerja RFECV (Sun <i>et al.</i> 2023)	9
3	Alur kerja ReliefF (Wu <i>et al.</i> 2020)	10
4	Ilustrasi cara kerja SVM (sumber: https://www.techslang.com/definition/what-is-an-svm/)	11
5	Ilustrasi sebuah Decision Tree (Saputra <i>et al.</i> 2023)	12
6	Alur Struktur Random Forest (Ahmad <i>et al.</i> 2018)	12
7	Ilustrasi cara kerja KNN (Sumber: https://www.linkedin.com/pulse/k-nearest-neighbors-knn-algorithm-classification-nrxgf/)	13
8	Persebaran kelas pada <i>dataset</i>	14
9	Alur Penelitian	15
10	Hasil Klasifikasi Decision Tree	22
11	Hasil Klasifikasi SVM	23
12	Hasil Klasifikasi Random Forest	24