

PROPOSAL

Optimasi Prediksi Kanker Payudara Berbasis Biomarker Metabolik Menggunakan Algoritma Support Vector Machine (SVM) dengan Bayesian Optimization



Disusun Oleh :

**Azmi Ittaqi Hammami (241572010007)
Shifi Amalia Zein (241552010013)**

Dosen Pengampu : Hendri Kharisma S.Kom, M.T

Prodi Teknik Informatika dan Sistem Informasi

**Sekolah Tinggi Manajemen Informatika dan Komputer TAZKIA.
Jl. Raya Dramaga Blok Radar Baru No.8, RT.03/RW.03, Margajaya, Kec. Bogor Bar.,
Kota Bogor, Jawa Barat 16116, Indonesia**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker payudara merupakan salah satu penyakit paling umum dan penyebab utama kematian akibat kanker pada wanita di seluruh dunia. Deteksi dini menjadi faktor yang sangat penting dalam meningkatkan peluang kesembuhan pasien. Namun meskipun demikian, proses diagnosis klinis seringkali membutuhkan biaya tinggi dan waktu lama. Pemanfaatan *Machine Learning* pada data biomarker darah (seperti Glukosa, Resistin, dll) menawarkan alternatif deteksi dini yang menjanjikan.

Penelitian terdahulu oleh **Patrício et al. (2018)** menggunakan dataset *Breast Cancer Coimbra* menunjukkan bahwa algoritma **Support Vector Machine (SVM)** memiliki kinerja diagnostik terbaik dibandingkan algoritma lain untuk jenis data ini. Namun, penelitian tersebut masih menggunakan metode pencarian parameter konvensional (*Grid Search*) yang cenderung tidak efisien secara komputasi dan kaku dalam mengeksplorasi ruang parameter.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi kanker payudara yang lebih optimal dengan menerapkan **SVM berbasis Bayesian Optimization**. Teknik optimasi ini dipilih karena kemampuannya mencari kombinasi *hyperparameter* terbaik secara adaptif dan efisien, sehingga diharapkan dapat menghasilkan model yang tidak hanya akurat tetapi juga stabil (*robust*) terhadap variasi data.

1.2 Rumusan Masalah

1. Bagaimana menerapkan algoritma *Support Vector Machine* (SVM) yang dioptimasi menggunakan *Bayesian Optimization* untuk diagnosis kanker payudara?
2. Bagaimana tingkat stabilitas (*robustness*) dan performa model SVM hasil optimasi dalam memprediksi status pasien berdasarkan data klinis?

1.3 Tujuan Penelitian

1. Membangun model klasifikasi SVM dengan penalaan parameter otomatis menggunakan metode *Bayesian Optimization*.
2. Mengevaluasi kinerja model menggunakan metrik AUC, Akurasi, dan Sensitivitas, serta menguji stabilitas model melalui validasi berulang (*repeated validation*).

1.4 Manfaat Penelitian

1. Memberikan alternatif metode optimasi model *machine learning* yang lebih efisien untuk data medis.
2. Menyediakan model deteksi dini yang teruji stabilitasnya untuk membantu tenaga medis.

BAB II

DESKRIPSI DATASET

2.1 Sumber Dataset

Dataset diambil dari UCI Machine Learning Repository:

- **Judul:** *Breast Cancer Coimbra Dataset*
- **URL:** <https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbra>

2.2 Fitur-Fitur Data

Kolom	Deskripsi
Age	Usia pasien
BMI	Body Mass Index
Glucose	Kadar glukosa darah
Insulin	Kadar insulin
HOMA	Homeostatic Model Assessment (indikator resistensi insulin)
Leptin	Hormon leptin
Adiponectin	Hormon adiponektin
Resistin	Protein resistin
MCP.1	Monocyte chemoattractant protein-1
Classification	Target class (0 = sehat, 1 = kanker payudara)

2.3 Preprocessing dan Kriteria Eksklusi

Sesuai dengan protokol medis yang dirujuk dari Patrício et al. (2018), dataset ini akan melalui tahap pembersihan data, di mana sampel dengan Indeks Massa Tubuh (BMI) $> 40 \text{ kg/m}^2$ akan dieksklusi untuk menjaga validitas biomarker metabolik (Resistin/Leptin) yang sensitif terhadap obesitas ekstrem.

2.4 Ukuran Dataset

- Jumlah data: 116 sampel
- Jumlah fitur: 10 (termasuk target class)
- Format file: **.CSV**
- Sumber: UCI Machine Learning Repository (2018)

BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama:

1. **Pengumpulan Data:**
Mengunduh dataset Breast Cancer Coimbra dari UCI Repository.
2. **Preprocessing Data:**
 - Pembersihan data (Filter BMI ≤ 40).
 - Standarisasi fitur (*Standard Scaler*) untuk menyamakan skala data biomarker.
 - Pembagian data latih dan uji.
3. **Pemodelan (Modeling):**
 - Menggunakan algoritma **Support Vector Machine (SVM)** dengan Kernel RBF (Radial Basis Function).
 - Menerapkan **Bayesian Optimization** (menggunakan library *Optuna*) untuk mencari *hyperparameter* optimal (C , γ , kernel).
4. **Evaluasi Model:**
 - Metrik Utama: *Area Under the Curve* (AUC) dan *Confusion Matrix*.
 - Uji Stabilitas: Melakukan pengujian berulang (*Repeated Random Sub-sampling*) sebanyak 20 iterasi untuk memastikan model tidak bias.

3.2 Alur Penelitian

Alur kerja penelitian ini dapat digambarkan sebagai berikut:

1. Input Dataset
2. Filter BMI & Scaling
3. Optimasi Parameter (Bayesian Opt)
4. Training SVM Terbaik
5. Validasi Robustness (20x)
6. Hasil & Analisis

BAB IV

KESIMPULAN DAN RENCANA

4.1 Hipotesis

Berdasarkan studi literatur, algoritma SVM yang dioptimasi dengan *Bayesian Optimization* diharapkan mampu mencapai performa yang kompetitif dengan metode konvensional, namun dengan efisiensi pencarian parameter yang lebih baik dan tingkat kesalahan (False Negative) yang terkendali.

4.2 Rencana Pengujian

Pengujian akan difokuskan pada pengukuran nilai AUC rata-rata dari berbagai skenario pembagian data acak untuk membuktikan bahwa model dapat diandalkan (*robust*) pada data pasien baru.

BAB V

TINJAUAN PUSTAKA

Patrício, M. et al. (2018). "Using resistin, glucose, age and BMI to predict the presence of breast cancer." Studi ini menjadi landasan utama pemilihan algoritma dalam proposal ini. Patrício membuktikan bahwa kombinasi 4 fitur (Glukosa, Resistin, Usia, BMI) memberikan hasil terbaik saat dimodelkan dengan **SVM** (AUC 0.87-0.91). Penelitian ini akan memperluas temuan tersebut dengan menerapkan teknik optimasi modern (*Bayesian Optimization*) pada algoritma SVM yang sama.