

NO: 01022479/INF/2024

**EVALUASI ALGORITMA *MACHINE LEARNING* DALAM MENENTUKAN
DIAGNOSIS DAN OBAT BERDASARKAN GEJALA PASIEN**

SKRIPSI

Diajukan untuk memenuhi persyaratan penyelesaian program S-1
Program Studi Informatika Fakultas Teknik Industri
Universitas Kristen Petra

Oleh :
Vincentius Andy
NRP: C14200191

PROGRAM STUDI INFORMATIKA



**FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS KRISTEN PETRA
SURABAYA
2024**

LEMBAR PENGESAHAN

SKRIPSI

**EVALUASI ALGORITMA *MACHINE LEARNING* DALAM MENENTUKAN
DIAGNOSIS DAN OBAT BERDASARKAN GEJALA PASIEN**

Oleh:

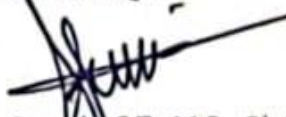
Vincentius Andy C14200191

Diterima oleh:

Program Studi Informatika Fakultas Teknologi Industri
Universitas Kristen Petra

Surabaya, 3 Juli 2024

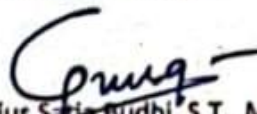
Pembimbing 1



Stephanus A. Ananda, S.T., M.Sc. Ph.D.

NIP: 93-026

Ketua Tim Penguji



Dr. Gregorius Satria Budhi, S.T., M.T.

NIP: 02-030

Ketua Program Studi



Edi Wibowo, S.T., M.T., Ph.D.

NIP: 00-003

**LEMBAR PERNYATAAN PERSETUJUAN
UNTUK KEPENTINGAN AKADEMIK**

Sebagai mahasiswa Universitas Kristen Petra, yang bertanda tangan dibawah ini saya:

Nama : Vincentius Andy

NRP : C14200191

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Petra Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul “EVALUASI ALGORITMA *MACHINE LEARNING* DALAM MENENTUKAN DIAGNOSIS DAN OBAT BERDASARKAN GEJALA PASIEN”. Dengan Hak Bebas Royalti Non-Eksklusif ini, Universitas Kristen Petra berhak menyimpan, mengalih-mediakan / formatkan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, menampilkan / mempublikasikannya di internet atau media lain untuk kepentingan akademik tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta.

Kami bersedia untuk menanggung secara pribadi, tanpa melibatkan Universitas Kristen Petra, segala bentuk tuntutan hukum yang timbul atau Pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan saya ini yang saya buat dengan sebenarnya.

Surabaya,



Yang menyatakan,
Vincentius Andy

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat rahmat dan anugerah-Nya, sehingga laporan penelitian skripsi dapat diselesaikan. Penelitian skripsi ini diajukan sebagai pemenuhan salah satu persyaratan dalam meraih gelar Sarjana Strata-1 di Jurusan Informatika, Fakultas Teknik Industri, Universitas Kristen Petra Surabaya.

Rasa terimakasih yang terdalam penulis sampaikan kepada semua pihak yang telah berperan besar dalam membantu penyelesaian skripsi ini, terutama :

1. Bapak Stephanus Ananda, selaku dosen pembimbing yang telah meluangkan waktu dalam memberikan bimbingan dan arahan dalam pengerjaan penelitian skripsi ini.
2. Kedua orang tua saya, dr. Rita Wey dan dr. Mahadi, yang tidak hanya membantu dengan memberikan dukungan moral dan emotional, tetapi juga membantu dalam proses pengerjaan selaku narasumber yang telah membantu proses validasi dari hasil penelitian skripsi.
3. Semua tim dari Apotek Sumber Sehati yang telah membantu dan mendukung dalam pengerjaan penelitian skripsi ini.
4. Segenap dosen dan staff pengajar di Jurusan Informatika Universitas Kristen Petra Surabaya
5. Semua pihak yang telah membantu baik secara langsung maupun tidak dalam penyelesaian penelitian skripsi ini yang tidak dapat disebutkan satu persatu

Penulis menyadari bahwa penulisan skripsi ini jauh dari sempurna, dan karena itu penulis mengharapkan bantuan dan saran yang membangun dari pembaca agar dapat mendukung perkembangan dan perbaikan dari pengerjaan skripsi dan penulis memohon maaf atas kekurangan dalam penulisan penelitian skripsi ini.

Surabaya

Penulis

ABSTRAK

Vincentius Andy

Skripsi

Evaluasi Algoritma *Machine Learning* Dalam Menentukan Diagnosis Dan Obat Berdasarkan Gejala Pasien

Penelitian ini bertujuan untuk menentukan algoritma machine learning terbaik dalam membantu dokter memberikan diagnosis dan rekomendasi obat berdasarkan gejala yang dialami pasien. Mengingat pentingnya diagnosis yang akurat dalam dunia medis, kesalahan diagnosis atau misdiagnosis dapat berdampak serius terhadap kesehatan pasien dan dapat menyebabkan konsekuensi hukum bagi tenaga medis. Dalam upaya ini, penelitian menggunakan beberapa algoritma machine learning, termasuk Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, dan Extreme Gradient Boosting, serta menerapkan metode ensemble seperti Bagging dan Voting untuk meningkatkan akurasi.

Data yang digunakan dalam penelitian ini diperoleh dari rekam medis pasien yang mencakup identitas pasien, gejala, diagnosis, dan obat yang diberikan. Dataset ini berasal dari hasil pemeriksaan yang dilakukan oleh dr. Rita Wey serta dari aplikasi web yang digunakan oleh klinik dan website PCare dari BPJS Kesehatan. Evaluasi akurasi model dilakukan dengan membandingkan hasil prediksi algoritma dengan diagnosis yang dibuat oleh tenaga kesehatan.

Hasil penelitian diharapkan dapat memberikan kontribusi signifikan dalam mengurangi kemungkinan kesalahan diagnosis dan meningkatkan akurasi dalam pemberian obat, sehingga membantu dokter dalam pengambilan keputusan medis yang lebih tepat dan efektif.

Kata Kunci :

machine learning, diagnosis, gejala, rekam medis, naïve bayes, decision tree, random forest, support vector machine, extreme gradient boosting, ensemble method.

ABSTRACT

Vincentius Andy

Undergraduate Thesis

Evaluation of Machine Learning Algorithms in Determining Diagnosis and Medication Based on Patient Symptoms

This research aims to determine the best machine learning algorithm to assist doctors in providing diagnoses and medication recommendations based on patients' symptoms. Given the critical importance of accurate diagnoses in the medical field, misdiagnosis can have serious impacts on patient health and can result in legal consequences for healthcare providers. To this end, the study employs several machine learning algorithms, including Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, and Extreme Gradient Boosting, and applies ensemble methods such as Bagging and Voting to improve accuracy.

The data used in this study is derived from patient medical records, including patient identity, symptoms, diagnoses, and medications given. This dataset comes from examinations conducted by dr. Rita Wey as well as from web applications used by clinics and the PCare website from BPJS Health. Model accuracy evaluation is conducted by comparing algorithm predictions with diagnoses made by healthcare professionals.

The results of this study are expected to make a significant contribution to reducing the likelihood of diagnostic errors and improving medication accuracy, thereby assisting doctors in making more precise and effective medical decisions.

Keywords

machine learning, diagnosis, symptoms, medical records, naïve bayes, decision tree, random forest, support vector machine, extreme gradient boosting, ensemble method.

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERNYATAAN PERSETUJUAN UNTUK KEPENTINGAN AKADEMIK.....	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xiii
DAFTAR SEGMENT PROGRAM.....	xiv
1. PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Tujuan Penelitian.....	2
1.4. Ruang Lingkup.....	3
2. LANDASAN TEORI.....	14
2.1. Tinjauan Pustaka.....	14
2.1.1. Diagnosis.....	14
2.1.2. Gejala.....	14
2.1.3. Naïve Bayes.....	14
2.1.4. Decision Tree.....	15
2.1.5. Random Forest.....	16
2.1.6. Support Vector Machine.....	17
2.1.7. Extreme Gradient Boosting.....	17
2.1.8. Evaluation Matrix.....	17
2.1.9. Ensemble Method.....	18
2.1.10. Voting.....	18
2.1.11. Bagging.....	19
2.1.12. Confusion Matrix.....	19
2.1.13. Dataset.....	19

2.1.14. Preprocessing.....	20
2.2. Tinjauan Studi.....	20
2.2.1. <i>Machine learning in medical applications: A review of state-of-the-art methods</i> (Shehab et al., 2022).....	20
2.2.2. <i>A machine learning based data modeling for medical diagnosis</i> (Mahoto et al., 2022).....	21
2.2.3. <i>Mobile Application Development for Disease Diagnosis based on Symptoms using Machine Learning Techniques</i> (Sridhar et al., 2023).....	22
3. ANALISIS DAN DESAIN SISTEM.....	27
3.1. Analisis Permasalahan.....	27
3.2. Analisis Kebutuhan.....	27
3.3. Pengumpulan Data.....	28
3.4. Pengolahan dan Analisis Data.....	30
3.5. Metode yang Diusulkan.....	30
3.6. Desain Model yang Dikembangkan.....	31
3.7. Proses Validasi.....	37
4. IMPLEMENTASI SISTEM.....	39
4.1. Implementasi dalam Pengerjaan Model.....	39
4.1.1. Implementasi Perangkat lunak.....	39
4.1.2. Implementasi Sistem.....	40
4.1.2.1. Data Loading.....	40
4.1.2.2. Preprocessing dan Normalisasi Data.....	41
4.1.2.3. Model Evaluation.....	41
4.1.2.4. Train-Test Splitting.....	42
4.1.2.5. Fitting dan Testing Model.....	42
4.1.2.6. Bagging Method.....	46
4.1.2.7. Voting Method.....	49
4.1.2.8. Data Cleaning untuk Prediksi Obat.....	51
4.1.2.9. Oversampling data obat.....	52
4.1.2.10. Train-Test Splitting.....	52
4.1.2.11. Testing Model.....	53
4.1.2.12. Bagging Method.....	53

4.1.2.13. Voting Method.....	53
4.2. Implementasi dalam Pengerjaan Aplikasi Berbasis Web untuk Validasi.....	53
4.2.1. Implementasi Perangkat Lunak.....	53
4.2.2. Implementasi Sistem.....	53
4.2.2.1. Implementasi dari <i>Back-End</i> Program.....	53
4.2.2.2. Implementasi dari <i>Front-End</i> Program.....	49
5. PENGUJIAN.....	64
5.1. Pengujian terhadap Hasil <i>Evaluation Matrix</i> Model.....	64
5.1.1. Diagnosis.....	64
5.1.1.1. Evaluasi pada <i>Base Model</i>	64
5.1.1.2. Evaluasi pada Bagging Method.....	69
5.1.1.3. Evaluasi pada Voting Method.....	73
5.1.2. Obat.....	76
5.1.2.1. Evaluasi pada <i>Base Model</i>	78
5.1.2.2. Evaluasi pada Bagging Method.....	83
5.1.2.3. Evaluasi pada Voting Method.....	88
5.2. Validasi Performa Algoritma Machine Learning oleh Pihak Tenaga Kesehatan.....	92
6. KESIMPULAN DAN SARAN.....	103
6.1. Kesimpulan.....	103
6.2. Saran.....	104
DAFTAR PUSTAKA.....	105

DAFTAR GAMBAR

Gambar 1.1. <i>Page UI</i> Hasil dari <i>Training</i> Algoritma.....	11
Gambar 1.2. <i>Page UI Input Data</i> dan Juga Hasil Prediksi.....	12
Gambar 2.1. Cara Kerja Algoritma Random Forest.....	17
Gambar 2.2. Contoh dari <i>Confusion Matrix</i>	19
Gambar 3.1. Contoh Data Dari Aplikasi Rekam Medis Sederhana.....	28
Gambar 3.2. Contoh Gambar Identitas Pasien dari Website BPJS.....	29
Gambar 3.3. Contoh Data Hasil Pemeriksaan Pasien dari Website BPJS.....	29
Gambar 3.4. <i>Flowchart</i> Tahapan Pengerjaan Penelitian Skripsi.....	31
Gambar 3.5. <i>Flowchart</i> cara kerja aplikasi berbasis web.....	32
Gambar 4.1. <i>Home Page</i> dari Aplikasi Berbasis Web.....	60
Gambar 4.2. <i>Patient Info Page</i> dari Aplikasi Berbasis Web untuk Menerima Input Karakteristik dan Gejala Pasien.....	61
Gambar 4.3. <i>Result Page</i> untuk Menampilkan Hasil dari Prediksi Berdasarkan Data yang Di-Inputkan.....	62
Gambar 4.4. <i>Alternative Medicine Page</i> untuk Menampilkan Obat Pengganti yang Dapat Diberikan Kepada Pasien.....	63
Gambar 5.1. <i>Evaluation Matrix</i> dari Model Naïve Bayes.....	64
Gambar 5.2. <i>Evaluation Matrix</i> dari Model Support Vector Machine.....	65
Gambar 5.3. <i>Evaluation Matrix</i> dari Model Random Forest.....	66
Gambar 5.4 <i>Evaluation Matrix</i> dari Model Decision Tree.....	67
Gambar 5.5. <i>Evaluation Matrix</i> dari Model Extreme Gradient Boosting.....	68
Gambar 5.6. <i>Evaluation Matrix</i> dari Model Naïve Bayes dengan Bagging Method.....	69
Gambar 5.7. <i>Evaluation Matrix</i> dari Model Support Vector Machine dengan Bagging Method.....	70
Gambar 5.8. <i>Evaluation Matrix</i> dari Model Random Forest dengan Bagging Method.....	71
Gambar 5.9. <i>Evaluation Matrix</i> dari Model Decision Tree dengan Bagging Method.....	72
Gambar 5.10. <i>Evaluation Matrix</i> dari Model Extreme Gradient Boosting dengan Bagging Method.....	73
Gambar 5.11. <i>Evaluation Matrix</i> dari Model Voting Method terhadap <i>Base Model</i>	74
Gambar 5.12. <i>Evaluation Matrix</i> dari Model Voting Method terhadap Bagging Method.....	75
Gambar 5.13. <i>Evaluation Matrix</i> dari Model Naïve Bayes untuk Prediksi Obat.....	79
Gambar 5.14. <i>Evaluation Matrix</i> dari Model Support Vector Machine untuk Prediksi Obat.....	80

Gambar 5.15. <i>Evaluation Matrix</i> dari Model Random Forest untuk Prediksi Obat.....	81
Gambar 5.16. <i>Evaluation Matrix</i> dari Model Decision Tree untuk Prediksi Obat.....	82
Gambar 5.17. <i>Evaluation Matrix</i> dari Model Extreme Gradient Boosting untuk Prediksi Obat....	83
Gambar 5.18. <i>Evaluation Matrix</i> dari Model Naïve Bayes dengan Bagging Method.....	84
Gambar 5.19. <i>Evaluation Matrix</i> dari Model Support Vector Machine dengan Bagging Method.....	85
Gambar 5.20. <i>Evaluation Matrix</i> dari Model Random Forest dengan Bagging Method.....	86
Gambar 5.21. <i>Evaluation Matrix</i> dari Model Decision Tree dengan Bagging Method.....	87
Gambar 5.22. <i>Evaluation Matrix</i> dari Model Extreme Gradient Boosting dengan Bagging Method.....	88
Gambar 5.23. <i>Evaluation Matrix</i> dari Model Voting Method terhadap <i>Base Model</i>	89
Gambar 5.24. <i>Evaluation Matrix</i> dari Model Voting Method terhadap Bagging Method.....	90
Gambar 5.25. Contoh Data Kunjungan yang Disimpan di Aplikasi Rekam Medis Sederhana.....	92
Gambar 5.26. Prediksi diagnosis A09 berdasarkan tenaga kesehatan.....	93
Gambar 5.27. Prediksi diagnosis A09 berdasarkan model Machine Learning.....	94
Gambar 5.28. Prediksi diagnosis K30 berdasarkan tenaga kesehatan.....	94
Gambar 5.29. Prediksi diagnosis K30 berdasarkan model Machine Learning.....	95
Gambar 5.30. Prediksi diagnosis m79.1 berdasarkan tenaga kesehatan.....	95
Gambar 5.31. Prediksi diagnosis m79.1 berdasarkan model Machine Learning.....	96
Gambar 5.32. Prediksi diagnosis I23.9 berdasarkan tenaga kesehatan.....	96
Gambar 5.33. Prediksi diagnosis I23.9 berdasarkan model Machine Learning.....	96
Gambar 5.34. Prediksi diagnosis E11 berdasarkan tenaga kesehatan.....	97
Gambar 5.35. Prediksi diagnosis E11 berdasarkan model Machine Learning.....	97
Gambar 5.36. Prediksi diagnosis I10 berdasarkan tenaga kesehatan.....	98
Gambar 5.37. Prediksi diagnosis I10 berdasarkan model Machine Learning.....	98
Gambar 5.38. Prediksi diagnosis R51 berdasarkan tenaga kesehatan.....	98
Gambar 5.39. Prediksi diagnosis R51 berdasarkan model Machine Learning.....	99
Gambar 5.40. Prediksi diagnosis J45 berdasarkan tenaga kesehatan.....	99
Gambar 5.41. Prediksi diagnosis J45 berdasarkan model Machine Learning.....	100
Gambar 5.42. Prediksi diagnosis j06.9 berdasarkan tenaga kesehatan.....	100
Gambar 5.43. Prediksi diagnosis j06.9 berdasarkan model Machine Learning.....	101
Gambar 5.44. Prediksi diagnosis R50.9 berdasarkan tenaga kesehatan.....	101

Gambar 5.45. Prediksi diagnosis R50.9 berdasarkan model Machine Learning.....	101
---	-----

DAFTAR TABEL

Tabel 2.1. Tabel Perbandingan State-Of-The-Art dari Penelitian Sebelumnya.....	23
Tabel 3.1. Tab Tabel parameter SVM yang diujikan.....	33
Tabel 3.2. Tab Tabel parameter Random Forest yang diujikan.....	34
Tabel 3.3. Tab Tabel parameter Decision Tree yang diujikan.....	34
Tabel 3.4. Tab Tabel parameter Naïve Bayes yang diujikan.....	35
Tabel 3.5. Tab Tabel parameter XGBoost yang diujikan.....	35
Tabel 3.6. Tab Tabel parameter Bagging Method yang diujikan.....	36
Tabel 3.7. Tab Tabel parameter Voting Method yang diujikan.....	37
Tabel 4.1. Tabel Penggunaan <i>Library</i> dan Masing – Masing Kegunaanya dalam Pengerjaan Model.....	39
Tabel 5.1. Tabel Hasil Evaluation Matrix dari Model dengan Parameter dan durasi <i>fitting</i> dari prediksi diagnosis	75
Tabel 5.2. Tabel Contoh Data dari Prediksi Obat.....	77
Tabel 5.3. Tabel Contoh Data Setelah Proses <i>Explode</i>	77
Tabel 5.4. Tabel Contoh Data Setelah Proses <i>Replacement</i>	78
Tabel 5.5. Tabel Hasil Evaluation Matrix dari Model dengan Parameter dan durasi <i>fitting</i> dari prediksi obat	90

DAFTAR SEGMENT PROGRAM

Segmen 4.1. Menghubungkan Google Colab dengan Dataset dari Google Drive.....	40
Segmen 4.2. Pengambilan Dataset yang Dibaca Menggunakan <i>Library</i> Pandas.....	40
Segmen 4.3. Proses Normalisasi Data.....	41
Segmen 4.4. <i>Function</i> Pengambilan Nilai <i>Evaluation Matrix</i>	41
Segmen 4.5. <i>Function</i> untuk Visualisasi <i>Confusion Matrix</i>	41
Segmen 4.6. <i>Train-Test Splitting</i>	42
Segmen 4.7. Deklarasi model dan parameter dari <i>base model</i>	43
Segmen 4.8. <i>Function</i> untuk melakukan <i>hyperparameter tuning</i>	43
Segmen 4.9. Deklarasi model dan parameter dari <i>base model</i>	44
Segmen 4.10. Deklarasi ulang model dengan parameter terbaik	44
Segmen 4.11. <i>Fitting Base</i> Model Naïve Bayes.....	45
Segmen 4.12. <i>Fitting Base</i> Model Support Vector Machine.....	45
Segmen 4.13. <i>Fitting Base</i> Model Random Forest.....	45
Segmen 4.14. <i>Fitting Base</i> Model Decision Tree.....	45
Segmen 4.15. <i>Fitting Base</i> Model Extreme Gradient Boosting	45
Segmen 4.16. <i>Function</i> penerapan <i>hyperparameter tuning</i> pada Bagging Method.....	46
Segmen 4.17. Proses <i>looping</i> dari <i>tuning</i> Bagging Method.....	47
Segmen 4.18. Deklarasi ulang dari model dengan Bagging Method	47
Segmen 4.19. <i>Fitting</i> Bagging Method pada Model Naïve Bayes.....	48
Segmen 4.20. <i>Fitting</i> Bagging Method pada Model Support Vector Machine.....	48
Segmen 4.21. <i>Fitting</i> Bagging Method pada Model Random Forest.....	48
Segmen 4.22. <i>Fitting</i> Bagging Method pada Model Decision Tree.....	48
Segmen 4.23. <i>Fitting</i> Bagging Method pada Model Extreme Gradient Boosting.....	48
Segmen 4.24. <i>Function</i> untuk melakukan <i>hyperparameter tuning</i> pada Voting Method.....	49
Segmen 4.25. Proses pengujian dari Voting Method dengan parameter yang diujikan.....	50
Segmen 4.26. Deklarasi ulang Voting Method dengan parameter terbaik.....	50
Segmen 4.27. <i>Fitting</i> Voting Method pada <i>Base Model</i>	51
Segmen 4.28. <i>Fitting</i> Voting Method pada Model dengan Bagging Classifier.....	51
Segmen 4.29. <i>Exploding</i> Data Berdasarkan Kolom Obat.....	51
Segmen 4.30. <i>Cleaning</i> Data Obat.....	51

Segmen 4.31. Penggantian Label Obat Sesuai dengan Data Wawancara.....	52
Segmen 4.32. <i>Oversampling</i> Data.....	52
Segmen 4.33. <i>Train-Test Splitting</i>	52
Segmen 4.34. <i>Framework</i> dari <i>Library</i> Flask.....	53
Segmen 4.35. <i>Function convert value</i>	53
Segmen 4.36. <i>Label Mapping</i> untuk Proses <i>Decoding</i> Hasil Prediksi.....	54
Segmen 4.37. <i>Loading</i> Model Kedalam Aplikasi Berbasis Web.....	55
Segmen 4.38. <i>Route</i> dalam Flask <i>Framework</i> Menuju <i>Home Page</i>	55
Segmen 4.39. <i>Route</i> dalam Flask <i>Framework</i> Menuju <i>Patient Info Page</i>	57
Segmen 4.40. Proses Prediksi Diagnosis.....	57
Segmen 4.41. Pengambilan Diagnosis Kedalam Data Prediksi Obat.....	58
Segmen 4.42. Proses Prediksi Obat.....	68