

**ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE
BAYES, DAN RANDOM FOREST DALAM KLASIFIKASI
PENERIMA VAKSIN INFLUENZA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DZAKY HAMID AL AZIZ

22.12.2587

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE
BAYES, DAN RANDOM FOREST DALAM KLASIFIKASI
PENERIMA VAKSIN INFLUENZA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

DZAKY HAMID AL AZIZ

22.12.2587

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE
BAYES, DAN RANDOM FOREST DALAM KLASIFIKASI
PENERIMA VAKSIN INFLUENZA**

yang disusun dan diajukan oleh

Dzaky Hamid Al Aziz

22.12.2587

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Februari 2026

Dosen Pembimbing,



Yoga Pristyanto, S.Kom., M.Eng.

NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE
BAYES, DAN RANDOM FOREST DALAM KLASIFIKASI
PENERIMA VAKSIN INFLUENZA**

yang disusun dan diajukan oleh

Dzaky Hamid Al Aziz

22.12.2587

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bety Wulan Sari, S.Kom., M.Kom.
NIK. 190302254

Sharazita Dyah Anggita, S.Kom., M.Kom.
NIK. 190302285

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dzaky Hamid Al Aziz
NIM : 22.12.2587

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE BAYES, DAN
RANDOM FOREST DALAM KLASIFIKASI PENERIMA VAKSIN
INFLUENZA**

Dosen Pembimbing : Yoga Pristyanto, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 Februari 2026

Yang Menyatakan,



Dzaky Hamid Al Aziz

HALAMAN PERSEMBAHAN

Dengan mengucapkan puji syukur kehadiran Allah SWT atas segala rahmat, karunia, dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi yang berjudul **“ANALISIS PERBANDINGAN ALGORITMA SVM, NAIVE BAYES, DAN RANDOM FOREST DALAM KLASIFIKASI PENERIMA VAKSIN INFLUENZA”**, dengan penuh rasa syukur dan kerendahan hati, karya sederhana ini saya persembahkan kepada:

1. Kedua orang tua tercinta yang telah menjadi sumber kekuatan terbesar dalam hidup saya, yang tidak pernah lelah memberikan dukungan, doa yang tulus, serta pengorbanan yang tak terhingga demi masa depan saya, sehingga saya dapat berada di titik ini.
2. Keluarga dan saudara-saudara yang senantiasa hadir memberikan dukungan moral, semangat, serta motivasi dalam setiap proses yang saya lalui, baik dalam suka maupun duka selama perjalanan menyelesaikan skripsi ini.
3. Dosen pembimbing Yoga Pristyanto, S.Kom., M.Eng. yang dengan penuh kesabaran telah membimbing, mengarahkan, serta memberikan ilmu dan masukan yang sangat berarti, sehingga saya mampu menyelesaikan skripsi ini dengan sebaik mungkin.
4. Teman-teman dan sahabat seperjuangan yang selalu ada dalam setiap proses, memberikan semangat, dukungan, serta kebersamaan yang tak ternilai, yang menjadikan perjalanan ini terasa lebih ringan dan penuh makna.
5. Almamater Universitas AMIKOM Yogyakarta yang telah menjadi tempat saya menimba ilmu, membentuk pola pikir, serta memberikan banyak pengalaman berharga yang akan saya bawa dalam perjalanan hidup ke depan.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Analisis Algoritma SVM, Naive Bayes, dan Random Forest dalam Klasifikasi Penerimaan Vaksin Influenza*” sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Anggit Dwi Hartanto, M.Kom. selaku Ketua Program Studi Sistem Informasi.
4. Yoga Pristyanto, S.Kom., M.Eng. selaku Dosen Pembimbing skripsi.
5. Kepada kedua orang tua yang telah memberikan dukungan dan doa.
6. Teman-teman seperjuangan saya yang selalu memberi semangat, kebersamaan, dan dorongan selama proses pengerjaan skripsi

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, sehingga kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan khususnya dalam machine learning dan analisis data.

Yogyakarta, 25 Februari 2026

Dzaky Hamid Al Aziz

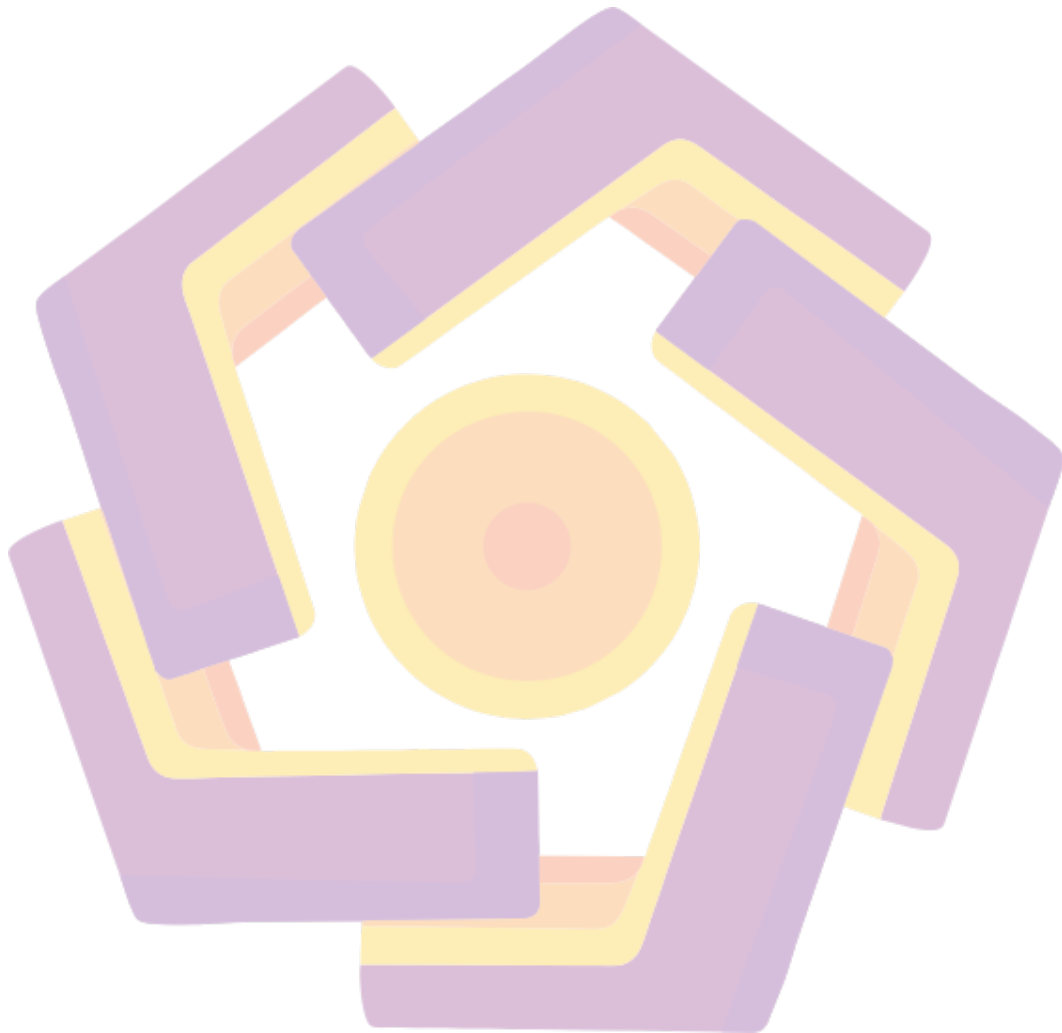
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH.....	xv
INTISARI	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3

1.6	Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA		6
2.1	Studi Literatur	6
2.2	Dasar Teori.....	17
2.2.1	Virus Influenza.....	17
2.2.2	Vaksin Influenza	18
2.2.3	Machine Learning	18
2.2.4	Data Mining	19
2.2.5	SMOTE (Synthetic Minority Oversampling Technique).....	19
2.2.6	CRISP-DM.....	20
2.2.7	Support Vector Machine	23
2.2.8	Naïve Bayes	24
2.2.9	Random Forest	25
2.2.10	Confusion Matrix	27
BAB III METODE PENELITIAN		31
3.1	Objek Penelitian.....	31
3.2	Alur Penelitian	31
3.2.1	Business Understanding.....	33
3.2.2	Data Understanding	33
3.2.3	Data Preparation.....	33
3.2.4	Modeling	34
3.2.5	Evaluation	34

3.2.6	Deployment.....	35
3.3	Alat dan Bahan.....	35
3.3.1	Deskripsi Dataset	35
3.3.2	Perangkat Keras	37
3.3.3	Perangkat Lunak.....	37
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Busines Understanding	39
4.2	Data Understanding	39
4.3	Data Preparation.....	47
4.3.1	Preprocessing Data.....	47
4.3.2	Data Split	50
4.3.3	SMOTE	50
4.4	Modeling.....	52
4.4.1	Hyperparameter Tuning	53
4.4.2	Pemodelan Data	54
4.5	Evaluasi Data	56
4.5.1	Confusion Matrix	56
4.5.2	Kurva ROC	69
4.4.3	Analisis Pebandingan.....	70
4.6	Deployment.....	74
BAB V PENUTUP		76
5.1	Kesimpulan	76

5.2	Saran	76
REFERENSI		78
LAMPIRAN.....		80



DAFTAR TABEL

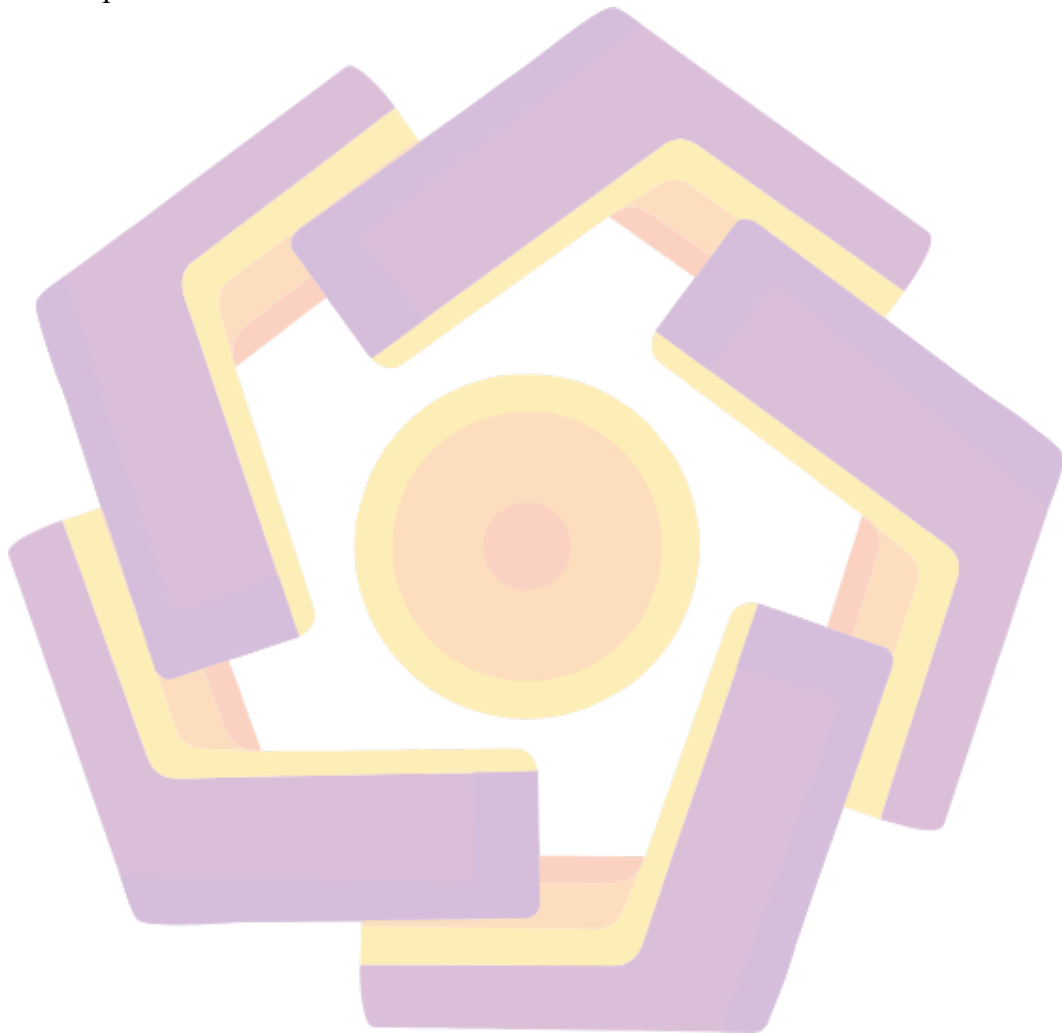
Tabel 2.1 Keaslian Penelitian	11
Tabel 2.2 Kelebihan dan Kekurangan Algoritma	26
Tabel 2.3 Confusion Matrix	27
Tabel 2.4 Nilai AUC	30
Tabel 3.1 Deskripsi Dataset	35
Tabel 3.2 Perangkat Keras	37
Tabel 3.3 Perangkat Lunak	37
Tabel 4.1 Fitur Numerikal	40
Tabel 4.2 Fitur Kategorikal	41
Tabel 4.3 Info Dataset	42
Tabel 4.4 Drop Kolom	47
Tabel 4.5 Hasil Hyperparameter Tuning	53
Tabel 4.6 Hasil Pemodelan	54
Tabel 4.7 Classification Report – SVM 10%	57
Tabel 4.8 Classification Report - Naive Bayes 10%	59
Tabel 4.9 Classification Report - Random Forest 10%	60
Tabel 4. 10 Classification Report – SVM 20%	61
Tabel 4. 11 Classification Report – Naive Bayes 20%	63
Tabel 4. 12 Classification Report – Random Forest 20%	64
Tabel 4. 13 Classification Report – SVM 30%	65
Tabel 4. 14 Classification Report – Naive Bayes 30%	67
Tabel 4. 15 Classification Report – Random Forest 30%	68
Tabel 4. 16 Perbandingan Algoritma	71

DAFTAR GAMBAR

Gambar 2.1 SMOTE	20
Gambar 2.2 Metode CRISP-DM	21
Gambar 4. 1 Deskripsi Dataset	39
Gambar 4. 2 Grafik Missing Value	44
Gambar 4.3 Kelas Target	45
Gambar 4. 4 Fitur Berpengaruh	46
Gambar 4.5 Seleksi Fitur dan Missing Value	49
Gambar 4.6 Fitur Kategorikal	50
Gambar 4.7 Sebelum Smote	51
Gambar 4.8 Setelah Smote	52
Gambar 4. 9 Confusion Matrix – SVM 10%	57
Gambar 4.10 Confusion Matrix - Naive Bayes 10%	58
Gambar 4.11 Confusion Matrix - Random Forest 10%	59
Gambar 4. 12 Confusion Matrix - SVM 20%	61
Gambar 4. 13 Confusion Matrix - Naive Bayes 20%	62
Gambar 4. 14 Confusion Matrix - Random Forest 20%	63
Gambar 4. 15 Confusion Matrix - SVM 30%	65
Gambar 4. 16 Confusion Matrix - Naive Bayes 30%	66
Gambar 4. 17 Confusion Matrix - Random Forest 30%	68
Gambar 4. 18 Kurva ROC	69
Gambar 4. 19 Visualisasi Perbandingan Akurasi	73
Gambar 4. 20 Hasil Deployment	74

DAFTAR LAMPIRAN

Lampiran 1. File Source Code dan Dataset	80
Lampiran 2. Surat SPD	80



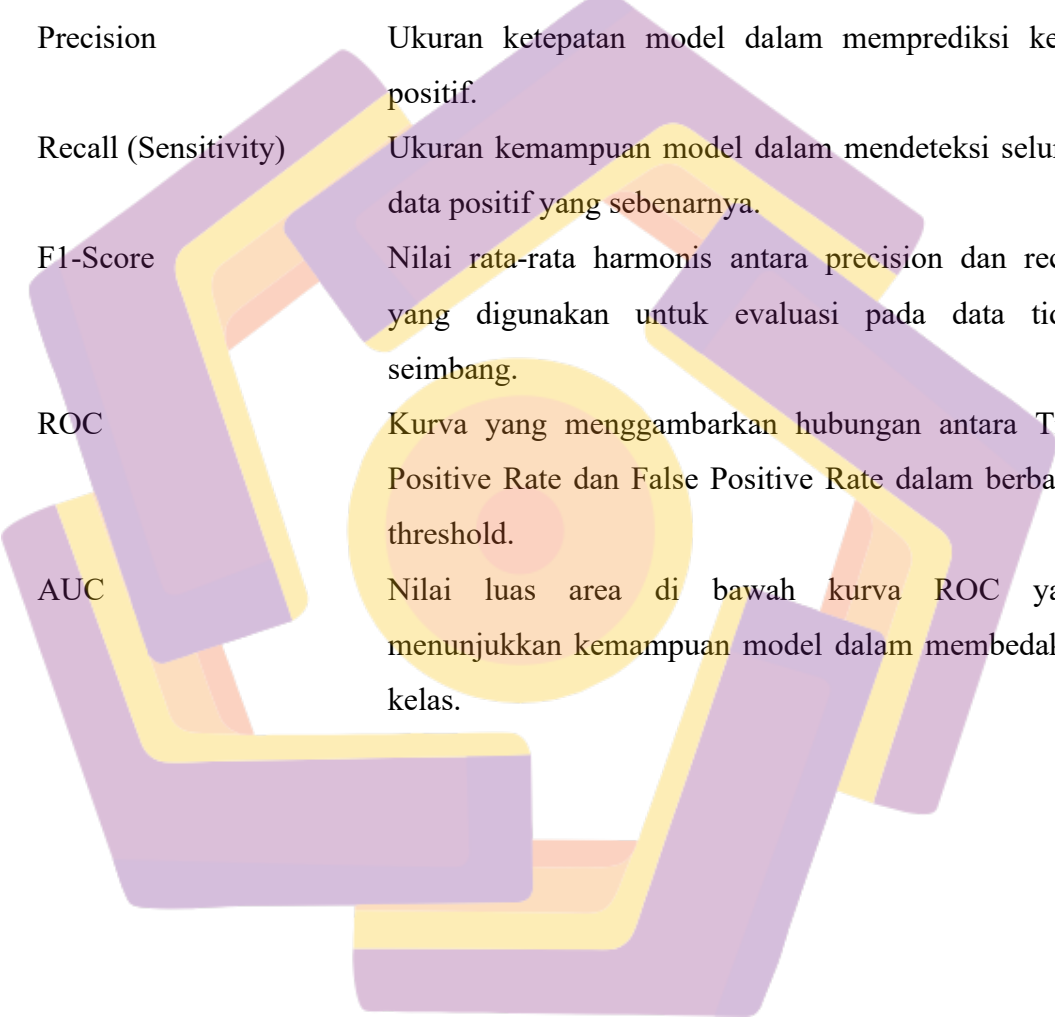
DAFTAR LAMBANG DAN SINGKATAN



SVM	Support Vector Machine
NB	Naïve Bayes
RF	Random Forest
SMOTE	Synthetic Minority Oversampling Technique
CRISP-DM	Cross-Industry Standard Process for Data Mining
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
ML	Machine Learning
EDA	Exploratory Data Analysis
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

DAFTAR ISTILAH

Virus Influenza	Virus yang menyerang sistem pernapasan dan dapat menyebabkan infeksi ringan hingga berat serta berpotensi menimbulkan komplikasi serius.
Vaksin Influenza	Produk digunakan untuk merangsang sistem imun agar membentuk perlindungan terhadap virus influenza.
Klasifikasi	Metode dalam data mining atau machine learning yang digunakan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan pola yang dipelajari dari data sebelumnya.
Machine Learning	Cabang dari kecerdasan buatan yang memungkinkan sistem komputer belajar dari data untuk membuat prediksi atau keputusan tanpa diprogram kembali
Dataset	Kumpulan data yang digunakan dalam penelitian untuk proses analisis dan pemodelan.
Fitur (Feature)	Atribut atau variabel input yang digunakan sebagai dasar dalam membangun model klasifikasi.
Data Mining	Proses menggali pola atau informasi tersembunyi dari data dalam jumlah besar menggunakan teknik statistik dan komputasi.
Imbalanced Class	Kondisi dataset di mana distribusi jumlah data antar kelas tidak seimbang.
Support Vector Machine	Algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal untuk memisahkan dua kelas dalam ruang fitur.
Naïve Bayes	Algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur.
Random Forest	Algoritma ensemble learning yang membangun



	banyak pohon keputusan dan menentukan hasil akhir melalui majority voting.
Hyperparameter Tuning	Proses optimasi parameter model untuk memperoleh performa terbaik.
Accuracy	Ukuran kinerja model yang menunjukkan proporsi prediksi yang benar terhadap seluruh data.
Precision	Ukuran ketepatan model dalam memprediksi kelas positif.
Recall (Sensitivity)	Ukuran kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya.
F1-Score	Nilai rata-rata harmonis antara precision dan recall yang digunakan untuk evaluasi pada data tidak seimbang.
ROC	Kurva yang menggambarkan hubungan antara True Positive Rate dan False Positive Rate dalam berbagai threshold.
AUC	Nilai luas area di bawah kurva ROC yang menunjukkan kemampuan model dalam membedakan kelas.

INTISARI

Teknologi dan akses kesehatan yang maju belum sepenuhnya mampu menghentikan penyebaran virus influenza yang masih menjadi ancaman kesehatan masyarakat di Indonesia akibat penyebarannya yang cepat. Meskipun vaksinasi merupakan tindakan pencegahan yang efektif, tingkat penerimaannya masih rendah karena faktor-faktor seperti kurangnya pemahaman, karakteristik individu, budaya, dan keterbatasan akses, yang mengakibatkan perlindungan kesehatan di masyarakat tidak merata. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Support Vector Machine (SVM), Naïve Bayes, dan Random Forest dalam mengklasifikasikan penerimaan vaksin influenza dengan menggunakan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining), yang mencakup tahapan dari business understanding hingga evaluation. Dengan menggunakan data sekunder dari Kaggle.com, hasil penelitian menunjukkan bahwa Random Forest mencapai kinerja terbaik dengan akurasi sebesar 85%, disusul oleh SVM sebesar 83% dan Naïve Bayes sebesar 70%. Hal ini membuktikan keunggulan Random Forest dalam menangani data yang kompleks dan tidak seimbang, dan temuan ini diharapkan dapat menjadi referensi ilmiah untuk pengembangan data yang lebih luas serta membantu proses pengambilan keputusan dan edukasi di bidang kesehatan masyarakat.

Kata kunci: Klasifikasi, Vaksin Influenza, Support Vector Machine, Naïve Bayes, Random Forest.

ABSTRACT

Advanced technology and healthcare access have not yet fully succeeded in halting the spread of the influenza virus, which remains a public health threat in Indonesia due to its rapid transmission. Although vaccination is an effective preventive measure, acceptance rates remain low due to factors such as a lack of understanding, individual characteristics, culture, and limited access, resulting in uneven health protection within the community. This study aims to analyze and compare the performance of Support Vector Machine (SVM), Naïve Bayes, and Random Forest algorithms in classifying influenza vaccine acceptance using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, encompassing stages from business understanding to evaluation. Utilizing secondary data from Kaggle.com, the results indicate that Random Forest achieved the best performance with an accuracy of 85%, followed by SVM at 83% and Naïve Bayes at 70%. This proves the superiority of Random Forest in handling complex and imbalanced data. Findings are expected to serve as a scientific reference for broader data development as well as to assist in decision-making and education processes in the field of public health.

Keyword: *Classification, Influenza Vaccine, Support Vector Machine, Naïve Bayes, Random Forest*