

**PERBANDINGAN KINERJA MODEL KLASIFIKASI
SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS
DALAM MENDIAGNOSIS ANEMIA**

BUYA AL-FARIZ



**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

@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 “Perbandingan Kinerja Model Klasifikasi *Support Vector Machine* dan *K-Nearest Neighbors* dalam Mendiagnosis Anemia” 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, Oktober 2024

Buya Al-Fariz
G5401201054

ABSTRAK

BUYA AL-FARIZ. Perbandingan Kinerja Model Klasifikasi *Support Vector Machine* dan *K-Nearest Neighbors* dalam Mendiagnosis Anemia. Dibimbing oleh MOCHAMAD TITO JULIANTO dan I WAYAN MANGKU.

Anemia merupakan masalah kesehatan yang ditandai dengan penurunan kadar hemoglobin di bawah batas normal. Pendekatan yang lebih baik dalam diagnosis anemia diperlukan untuk memastikan penanganan yang tepat. Penelitian ini bertujuan membandingkan kinerja beberapa model klasifikasi pada *machine learning* untuk mendiagnosis anemia menggunakan data dengan variabel *gender*, *hemoglobin*, *hematocrit*, *MCH*, *MCHC*, *MCV*, *RBC*, dan variabel anemia. Empat model yang digunakan, yaitu *support vector machine* (SVM) kernel linear, *support vector machine* (SVM) kernel *polynomial*, *support vector machine* (SVM) kernel *radial basis function*, dan *k-nearest neighbors* (KNN). Data diolah menggunakan metode *Mahalanobis distance* untuk penanganan pencilan, *min-max scaler* untuk transformasi data, dan *hyperparameters tuning* untuk mengoptimalkan kinerja model. Berdasarkan hasil penelitian, SVM kernel *polynomial* menunjukkan kinerja terbaik dari segi *accuracy*, *precision*, *recall*, dan *f1-score*. Namun, dari sisi efisiensi, KNN menjadi yang paling unggul. Sementara itu, SVM kernel linear menunjukkan keseimbangan terbaik antara kinerja dan efisiensi sehingga dapat dipertimbangkan sebagai alternatif yang optimal dalam diagnosis anemia.

Kata kunci: anemia, knn, *machine learning*, svm.

ABSTRACT

BUYA AL-FARIZ. Performance Comparison of Support Vector Machine and K-Nearest Neighbors Classification Models in Diagnosing Anemia. Supervised by MOCHAMAD TITO JULIANTO and I WAYAN MANGKU.

Anemia is an health problem characterized by a decrease in hemoglobin levels below normal limits. A better approach in the diagnosis of anemia is needed to ensure proper treatment. This study aims to compare the performance of several classification models in machine learning to diagnose anemia using data with gender, hemoglobin, hematocrit, MCH, MCHC, MCV, RBC, and anemia variables. Four models are used, namely support vector machine (SVM) linear kernel, support vector machine (SVM) polynomial kernel, support vector machine (SVM) radial basis function kernel, and k-nearest neighbors (KNN). Data are processed using the Mahalanobis distance method for outlier handling, min-max scaler for data transformation, and hyperparameters tuning to optimize model performance. Based on the results, SVM kernel polynomial showed the best performance in terms of accuracy, precision, recall, and f1-score. However, in terms of efficiency, KNN is the most superior. Meanwhile, linear kernel SVM shows the best balance between performance and efficiency so that it can be considered as an optimal alternative in anemia diagnosis.

Keywords: anemia, knn, machine learning, svm.



@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.

© Hak Cipta milik IPB, tahun 2024
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.

PERBANDINGAN KINERJA MODEL KLASIFIKASI SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS DALAM MENDIAGNOSIS ANEMIA

BUYA AL-FARIZ

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

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

IPB University

Penguji pada Ujian Skripsi:
Prof. Dr. Ir. Sri Nurdianti, M.Sc.



IPB University
— Bogor Indonesia —

- 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.

Perpustakaan IPB University

Judul Skripsi : Perbandingan Kinerja Model Klasifikasi *Support Vector Machine* dan *K-Nearest Neighbors* dalam Mendiagnosis Anemia

Nama : Buya Al-Fariz

NIM : G5401201054

Disetujui oleh

Pembimbing 1:

Mochamad Tito Julianto, S.Si., M.Kom.

Pembimbing 2:

Prof. Dr. Ir. I Wayan Mangku, M.Sc.

Diketahui oleh

Ketua Departemen Matematika:

Dr. Ir. Endar Hasafah Nugrahani, M.S.

NIP 196312281989032001

Tanggal Ujian: 5 September 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 Januari 2024 sampai bulan Agustus 2024 ini ialah klasifikasi data, dengan judul “Perbandingan Kinerja Model Klasifikasi *Support Vector Machine* dan *K-Nearest Neighbors* dalam Mendiagnosis Anemia”.

Terima kasih penulis ucapkan kepada seluruh pihak yang telah membantu dalam proses penelitian dan penyusunan karya ilmiah ini, antara lain kepada:

1. Kedua orang tua penulis, yaitu Bapak Imam Budoyo dan Ibu Sri Andarwati, nenek penulis, yaitu Sunarti, serta kakak dan adik penulis, yaitu Alam Priyo Nugroho, Laras Tunjung Irmada, Novita Duri, dan Miranda Ayu Artika yang selalu memberikan doa, kasih sayang, nasihat, dukungan penuh, dan menjadi tujuan dalam setiap langkah hidup penulis.
2. Bapak Mochamad Tito Julianto, S.Si., M.Kom. selaku dosen pembimbing tugas akhir 1 yang telah membimbing, membina, dan memberi arahan kepada penulis selama penyusunan karya ilmiah ini.
3. Bapak Prof. Dr. Ir. I Wayan Mangku, M.Sc. selaku dosen pembimbing tugas akhir 2 yang telah membimbing dan banyak memberikan masukan serta saran kepada penulis selama penyusunan karya ilmiah ini.
4. Bapak Drs. Prapto Tri Supriyo, M.Kom. selaku dosen pembimbing akademik yang telah mendampingi dan memberikan arahan selama masa perkuliahan.
5. Seluruh dosen dan staf Departemen Matematika IPB atas bimbingan, ilmu yang bermanfaat, bantuan, serta kebaikannya selama penulis menempuh perkuliahan.
6. Muhammad Faqih Murtaqi, Abdul Zumar Fauzan, dan Bagas Andhika Pramudya yang menemani, membantu, dan memberikan dukungan serta motivasi kepada penulis selama masa perkuliahan dan penyusunan karya ilmiah ini.
7. Hafidz Al Ubaidah, Noval, Syahrul, Muhammad Taufik Ikhwanudin, Sapto Mukti Handoyo, Hari Nurdianto, dan Muhammad Aryo Wicaksono sebagai teman dekat yang menemani dan memberikan dukungan kepada penulis selama masa perkuliahan dan penyusunan karya ilmiah ini.
8. Teman-teman Departemen Matematika 57 yang sudah menemani perjalanan empat tahun di IPB.

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

Bogor, Oktober 2024

Buya Al-Fariz



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Anemia	3
2.2 Pra-pemrosesan Data	3
2.3 <i>Machine Learning</i>	6
2.4 <i>Support Vector Machine</i>	7
2.5 <i>K-Nearest Neighbors</i>	11
2.6 <i>Model Performance</i>	13
2.7 <i>Receiver Operating Characteristic Curve (ROC Curve)</i>	14
III METODE	16
3.1 Sumber Data	16
3.2 Prosedur Analisis Data	18
3.3 Diagram Alir Penelitian	19
3.4 Perangkat Lunak	20
IV HASIL DAN PEMBAHASAN	21
4.1 Pra-pemrosesan Data	21
4.2 Implementasi dan Analisis Kinerja Model	28
4.3 Evaluasi Model	36
V SIMPULAN DAN SARAN	38
5.1 Simpulan	38
5.2 Saran	38
DAFTAR PUSTAKA	39
LAMPIRAN	42
RIWAYAT HIDUP	47



DAFTAR TABEL

1	Tabulasi setiap variabel dalam dataset	17
2	Deteksi <i>missing data</i> pada setiap variabel	21
3	Deteksi pencilan berdasarkan <i>mahalanobis distance</i>	25
4	Skor <i>accuracy</i> setiap alternatif nilai parameter <i>C</i>	28
5	Skor <i>accuracy</i> setiap kombinasi nilai parameter <i>C</i> dan <i>degree</i>	30
6	Skor <i>accuracy</i> setiap kombinasi nilai parameter <i>C</i> dan <i>gamma</i>	32
7	Skor <i>accuracy</i> setiap kombinasi nilai parameter <i>n_neighbors</i> dan <i>weights</i>	34
8	Nilai metrik evaluasi setiap model	36

DAFTAR GAMBAR

1	Ilustrasi <i>hyperplane</i> dan margin antara dua <i>class</i>	8
2	Ilustrasi <i>hyperplane</i> dan margin optimal antara dua <i>class</i> pada <i>support vector machine</i>	8
3	Pemetaan pada <i>support vector machine</i>	10
4	Ilustrasi <i>k-nearest neighbors</i>	11
5	<i>Confusion matrix</i>	13
6	ROC-AUC	15
7	Diagram alir penelitian	19
8	Proporsi tiap kelas variabel respons	22
9	Perbandingan anemia dan tidak anemia berdasarkan variabel gender	23
10	Visualisasi korelasi antarvariabel menggunakan <i>heatmap</i>	24
11	Visualisasi korelasi antara variabel <i>hemoglobin</i> dan HCT (<i>hematocrit</i>)	25
12	Perbandingan distribusi variabel <i>hemoglobin</i> sebelum dan sesudah transformasi	26
13	Perbandingan distribusi variabel HCT sebelum dan sesudah transformasi	26
14	Perbandingan distribusi variabel MCH sebelum dan sesudah transformasi	27
15	Perbandingan distribusi variabel MCHC sebelum dan sesudah transformasi	27
16	Perbandingan distribusi variabel MCV sebelum dan sesudah transformasi	27
17	Perbandingan distribusi variabel RBC sebelum dan sesudah transformasi	27
18	<i>Confusion matrix</i> model SVM kernel linear dengan <i>min-max scaler</i>	29
19	Hasil <i>Receiver Operating Characteristic</i> (ROC) model SVM kernel linear dengan <i>min-max scaler</i>	29
20	<i>Confusion matrix</i> model SVM kernel <i>polynomial</i> dengan <i>min-max scaler</i>	31
21	Hasil <i>Receiver Operating Characteristic</i> (ROC) model SVM kernel <i>polynomial</i> dengan <i>min-max scaler</i>	31

22	<i>Confusion matrix model SVM kernel radial basis function dengan min-max scaler</i>	33
23	Hasil <i>Receiver Operating Characteristic (ROC)</i> model SVM kernel <i>radial basis function</i> dengan <i>min-max scaler</i>	33
24	<i>Confusion matrix model KNN dengan min-max scaler</i>	35
25	Hasil <i>Receiver Operating Characteristic (ROC)</i> model KNN dengan <i>min-max scaler</i>	35

DAFTAR LAMPIRAN

1	Pengecekan <i>missing data</i>	43
2	Pengecekan pencilan	43
3	Pencarian parameter terbaik pada <i>support vector machine</i> kernel linear dengan <i>min-max scaler</i>	44
4	Pencarian parameter terbaik pada <i>support vector machine</i> kernel <i>polynomial</i> dengan <i>min-max scaler</i>	44
5	Pencarian parameter terbaik pada <i>support vector machine</i> kernel <i>radial basis function</i> dengan <i>min-max scaler</i>	44
6	Pencarian parameter terbaik pada <i>k-nearest neighbors</i> dengan <i>min-max scaler</i>	45
7	Evaluasi kinerja model <i>support vector machine</i> kernel linear menggunakan parameter terbaik	45
8	Evaluasi kinerja model <i>support vector machine</i> kernel <i>polynomial</i> menggunakan parameter terbaik	45
9	Evaluasi kinerja model <i>support vector machine</i> kernel <i>radial basis function</i> menggunakan parameter terbaik	46
10	Evaluasi kinerja model <i>k-nearest neighbors</i> menggunakan parameter terbaik	46