



**PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK PREDIKSI PENYAKIT TIROID**

SKRIPSI

MAULANA LUTHFI

1910511078

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
JAKARTA**

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA**

2023



**PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI
UNTUK PREDIKSI PENYAKIT TIROID**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Komputer**

MAULANA LUTHFI

1910511078

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
JAKARTA**

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA**

2023

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun ditujuk telah saya nyatakan dengan benar.

Nama : Maulana Luthfi

NIM : 1910511078

Program Studi : S1 Informatika

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 05 Juni 2023

Yang Menyatakan,



(Maulana Luthfi)

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini :

Nama : Maulana Luthfi
NIM : 1910511078
Fakultas : Ilmu Komputer
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Perbandingan Kinerja Algoritma Klasifikasi untuk Prediksi Penyakit Tiroid

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih-media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 17 Juli 2023

Yang menyatakan,



(Maulana Luthfi)


LEMBAR PENGESAHAN

LEMBAR PENGESAHAN


Tugas Akhir ini diajukan oleh:

Nama : Maulana Luthfi
NIM : 1910511078
Program Studi : SI Informatika
Judul Tugas Akhir : Perbandingan Kinerja Algoritma Klasifikasi untuk
Prediksi Penyakit Tiroid

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.


(Bayu Hananto, S.Kom. M.Kom.)

Penguji I


(Ika Nurlaili Isnainiyah, S.Kom. M.Sc.)


Penguji II


(Dr. Didit Widivanto, S.Kom. M.Si.)

Dosen Pembimbing



(Dr. Ermatita, M.Kom.)
Dekan Fakultas Ilmu Komputer


(Dr. Widva Cholil, M.I.T)
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 06 Juli 2023



PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI UNTUK PREDIKSI PENYAKIT TIROID

Maulana Luthfi

ABSTRAK

Tiroid merupakan salah satu bagian tubuh yang sangat penting bagi manusia. Gangguan tiroid seringkali sulit dideteksi karena memiliki gejala yang sangat mirip dengan penyakit lain. Akibatnya pasien seringkali tidak menyadari memiliki masalah pada kelenjar tiroid.

Penelitian ini bertujuan untuk melakukan prediksi penyakit tiroid menggunakan data penyakit tiroid dari Garavan Institute dan J. Ross Quinlan. Data tersebut akan diolah menggunakan proses data mining menggunakan metode klasifikasi.

Metode klasifikasi yang digunakan pada penelitian ini adalah algoritma K-Nearest Neighbor, Naïve Bayes, Decision Tree, dan Random Forest. Penelitian ini juga menggunakan teknik *undersampling* untuk mengatasi jumlah kelas yang tidak seimbang. Keluaran dari sistem adalah *performance measure* yang disajikan dalam bentuk persentase berdasarkan nilai – nilai pada *confusion matrix* dari setiap algoritma terhadap data penyakit tiroid.

Berdasarkan hasil yang diperoleh, kinerja algoritma Random Forest memperoleh hasil paling baik dibandingkan dengan algoritma lainnya. Random Forest mendapat persentase akurasi sebesar 95,16%, sedangkan KNN, Naïve Bayes, dan Decision Tree hanya mendapat persentase akurasi sebesar 87,10%, 93,55%, dan 90,32%. Namun jika dilihat dari waktu proses pelatihan dan pengujian model, Decision Tree memiliki waktu proses paling cepat dibandingkan dengan algoritma lainnya. Decision Tree membutuhkan waktu 0,0091 detik, sedangkan KNN, Naïve Bayes, dan Random Forest membutuhkan waktu 0,0273 detik, 0,0103 detik, dan 0,0162 detik.

Oleh karena itu jika ingin mendapatkan nilai yang maksimal, maka dapat menggunakan algoritma Random Forest. Namun jika ingin menggunakan waktu proses yang lebih cepat, maka dapat menggunakan algoritma Decision Tree untuk melakukan prediksi penderita penyakit tiroid.

Kata kunci : *Tiroid, Data Mining, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Random Forest, Undersampling, Confusion Matrix, Performance Measure*

COMPARISON OF CLASSIFICATION ALGORITHM PERFORMANCE FOR THYROID DISEASE PREDICTION

Maulana Luthfi

ABSTRACT

Thyroid is a very important part of the body for humans. Thyroid disorders are often difficult to detect because their symptoms are very similar to those of other diseases. As a result, patients often do not realize they have problems with the thyroid gland.

This study aims to predict thyroid disease using thyroid disease data from the Garavan Institute and J. Ross Quinlan. The data will be processed using the data mining process using the classification method.

The classification methods used in this study are the K-Nearest Neighbor, Naïve Bayes, Decision Tree, and Random Forest algorithms. This study also uses undersampling techniques to overcome the unequal number of classes. The output of the system is a performance measure that is presented in the form of a proportion based on the values in the confusion matrix of each algorithm against thyroid disease data.

Based on the results obtained, the performance of the Random Forest algorithm obtained the best results compared to other algorithms. Random Forest got an accuracy proportion of 95.16%, while KNN, Naïve Bayes, and Decision Tree only got an accuracy proportion of 87.10%, 93.55% and 90.32%. However, when viewed from the training process time and the testing model, Decision Tree has the fastest processing time compared to other algorithms. Decision Tree takes 0.0091 seconds, while KNN, Naïve Bayes, and Random Forest takes 0.0273 seconds, 0.0103 seconds and 0.0162 seconds.

Therefore if you want to get the maximum value, you can use the Random Forest algorithm. However, if you want to use a faster processing time, you can use the Decision Tree algorithm to predict thyroid disease sufferers.

Keywords : *Thyroid, Data Mining, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Random Forest, Undersampling, Confusion Matrix, Performance Measure*

KATA PENGANTAR

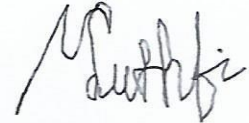
Puji dan syukur kehadiran Tuhan Yang Maha Esa atas berkat dan Rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul “Perbandingan Kinerja Algoritma Klasifikasi untuk Prediksi Penyakit Tiroid” dengan baik. Tugas akhir ini ditulis sebagai salah satu syarat untuk memperoleh gelar sarjana program studi Informatika, Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Penulis juga mengucapkan terima kasih kepada seluruh pihak yang telah membantu dan memberi semangat dalam menyelesaikan tugas akhir ini sehingga pengerjaan tugas akhir ini dapat dikerjakan dengan baik. Penulis mengucapkan terima kasih banyak kepada:

1. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer UPN Veteran Jakarta
2. Ibu Dr. Widya Cholil, M.I.T selaku Kaprodi Informatika UPN Veteran Jakarta
3. Bapak Dr. Didit Widiyanto, S.Kom, M.Si selaku Dosen Pembimbing, yang telah memberikan bimbingan dan masukan sehingga penulis dapat menyelesaikan tugas akhir ini.
4. Bapak Hamonangan Kinantan Prabu, S.T. selaku Dosen Pembimbing Akademik
5. Seluruh dosen Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta yang telah memberikan bekal ilmu selama saya menempuh studi.
6. Teristimewa kepada keluarga besar penulis, terutama pada Ibu yang sudah memberikan kasih dan sayang kepada penulis mulai dari kecil sampai sekarang, Tanpa cinta dari keluarga mungkin skripsi ini tidak dapat diselesaikan.
7. Para sahabat pada kelompok Pejuang yang selalu memberikan dukungan dan semangat.

Penulis menyadari pada tugas akhir ini sangat dimungkinkan masih banyak kekurangan yang harus diperbaiki. Segala bentuk kritik dan saran akan dengan senang hati diterima dan diharapkan dapat membantu memperbaiki tugas akhir ini. Semoga tugas akhir ini dapat berguna dan bermanfaat bagi semua pihak. Terima kasih.

Jakarta, 05 Juni 2023

A handwritten signature in black ink, appearing to read 'Maulana Luthfi', written in a cursive style.

Maulana Luthfi

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT.....	vii
KATA PENGANTAR.....	viii
Daftar Gambar	xiii
Daftar Tabel.....	xiv
Daftar Rumus	xv
Daftar Lampiran	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup.....	2
1.4 Tujuan dan Manfaat Penelitian	3
1.4.1 Tujuan Penelitian.....	3
1.4.2 Manfaat Penelitian.....	3
1.5 Luaran.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Tiroid.....	5
2.2.1 Jenis Penyakit Tiroid.....	5
2.2.2 Gejala Penyakit Tiroid	6
2.2.3 Penyebab Penyakit Tiroid	6
2.2 Data Mining.....	6

2.3 <i>Undersampling</i>	7
2.4 K-Nearest Neighbor	9
2.5 Naïve Bayes.....	9
2.6 Decision Tree	10
2.7 Random Forest	11
2.8 <i>Confusion Matrix</i>	12
2.9 Kajian Pustaka.....	14
BAB III METODELOGI PENELITIAN.....	17
3.1 Identifikasi Masalah	17
3.2 Studi Literatur	18
3.3 Akuisisi Data	18
3.4 Dataset.....	18
3.5 Data Preprocessing	22
3.6 <i>Undersampling</i>	22
3.7 Pemisahan Data	22
3.8 Pelatihan dan Pengujian Model.....	23
3.9 Analisis dan Kesimpulan.....	25
3.10 Alat Bantu Penelitian	25
3.11.1 Perangkat Keras.....	25
3.11.2 Perangkat Lunak.....	25
3.11 Jadwal Penelitian.....	26
BAB IV HASIL DAN PEMBAHASAN	27
4.1 <i>Data Preprocessing</i>	27
4.2.1 Menghapus Duplikat	27
4.2.2 <i>Data Cleaning</i>	27
4.2 <i>Undersampling</i>	28

4.3 Data <i>Splitting</i>	29
4.4 Algoritma K-Nearest Neighbor.....	29
4.5 Algoritma Naïve Bayes	36
4.6 Algoritma Decision Tree.....	41
4.7 Algoritma Random Forest.....	48
4.8 Perbandingan Hasil Klasifikasi Algoritma.....	55
4.8.1. Perbandingan Hasil Klasifikasi Berdasarkan Waktu Proses	55
4.8.2. Perbandingan Hasil Klasifikasi Berdasarkan Nilai Akurasi	56
4.8.3. Perbandingan Hasil Klasifikasi Berdasarkan Kinerja	56
BAB V PENUTUP.....	58
5.1 Kesimpulan.....	58
5.2 Saran.....	59
DAFTAR PUSTAKA	60
RIWAYAT HIDUP	61
Lampiran	62

Daftar Gambar

Gambar 3.1. Tahapan Penelitian.....	17
Gambar 4.1. Hasil Akurasi KNN	30
Gambar 4.2. Confusion Matrix KNN	30
Gambar 4.3. Confusion Matrix Naïve Bayes	36
Gambar 4.4. Hasil Tree Decision Tree	42
Gambar 4.5. Confusion Matrix Decision Tree	43
Gambar 4.6. Hasil Akurasi Random Forest.....	48
Gambar 4.7. Waktu yang Dibutuhkan Nilai N	49
Gambar 4.8. Confusion Matrix Random Forest	50
Gambar 4.9. Waktu Proses Setiap Algoritma.....	55
Gambar 4.10. Hasil Akurasi Setiap Algoritma.....	56

Daftar Tabel

Tabel 3.1. Parameter.....	19
Tabel 3.2. Jadwal Penelitian.....	26
Tabel 4.1. Performance Measure KNN	33
Tabel 4.2. Performance Measure Naïve Bayes	38
Tabel 4.3. Performance Measure Decision Tree	45
Tabel 4.4. Waktu yang Dibutuhkan Nilai N.....	49
Tabel 4.5. Performance Measure Random Forest	52
Tabel 4.6. Hasil Performance Measure Setiap Algoritma.....	56

Daftar Rumus

Rumus 2.1. Perhitungan Eucladian Distance	9
Rumus 2.2. Perhitungan Naïve Bayes	10
Rumus 2.3. Perhitungan Gini Index Decision Tree	11
Rumus 2.4. Perhitungan Accuracy	13
Rumus 2.5. Perhitungan Precision	13
Rumus 2.6. Perhitungan Recall	13
Rumus 2.7. Perhitungan F-1 Score	14

Daftar Lampiran

Lampiran 1	62
Lampiran 2	69