IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI TINGKAT RISIKO MENGIDAP KANKER PARU-PARU

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Nehemia Gueldi
00000043202

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI TINGKAT RISIKO MENGIDAP KANKER PARU-PARU

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Nehemia Gueldi
00000043202

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Nehemia Gueldi
Nomor Induk Mahasiswa	:	00000043202
Program Studi	:	Informatika

Skripsi dengan judul:

Implementasi Algoritma Random Forest Untuk Klasifikasi Tingkat Risiko Mengidap Kanker Paru-Paru

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 15 Mei 2024

(Nehemia Gueldi)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI TINGKAT RISIKO MENGIDAP KANKER PARU-PARU

oleh

Nama	:	Nehemia Gueldi
NIM	:	00000043202
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 07 Juni 2024
Pukul 10.00 s/d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Adhi Kusnadi, S.T, M.Si.)
NIDN: 0303037304

Penguji

(Dr. Maria Irmina Prasetiyowati, S.Kom., M.T.)
NIDN: 0725057201

Pembimbing

(Dennis Gunawan, S.Kom., M.Sc.)
NIDN: 0320059001

PJS Ketua Program Studi Informatika,

Dr. Eng. Niki Prastomo, S.T., M.Sc.
NIDN: 0419128203

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Nehemia Gueldi
NIM	:	00000043202
Program Studi	:	Informatika
Jenjang	:	S1
Jenis Karya	:	Skripsi

Menyatakan dengan sesungguhnya bahwa:

☐ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

☑ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 3 Juni 2024

Yang menyatakan

Nehemia Gueldi

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

”A good name is to be more desired than great wealth, Favor is better than silver and gold.”

Proverbs 22:1 (NASB)

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma Random Forest Untuk Klasifikasi Tingkat Risiko Mengidap Kanker Paru-Paru dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Pjs. Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Orang Tua, dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 3 Juni 2024

Nehemia Gueldi

IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI TINGKAT RISIKO MENGIDAP KANKER PARU-PARU

Nehemia Gueldi

ABSTRAK

Kanker paru-paru merupakan salah satu penyakit yang menjadi fokus utama dalam bidang kesehatan global karena memiliki jumlah kasus kematian yang tinggi. Meskipun jumlah kasus kanker payudara pada wanita melampaui kanker paru-paru, kanker paru-paru tetap menjadi penyebab utama kematian. Faktor risiko utama kanker paru-paru adalah merokok dan paparan polusi udara. Dengan membangun sejumlah pohon keputusan secara terpisah dan menggabungkan hasilnya, algoritma *Random Forest* dapat melakukan klasifikasi yang akurat dan mencegah *overfitting*. Pengujian dilakukan menggunakan *dataset Cancer Patient Datasets* untuk mengukur *accuracy*, *precision*, *recall*, dan *F1 score* yang diperoleh. Hasil yang didapatkan dalam penelitian ini adalah algoritma *Random Forest* berhasil diimplementasikan untuk klasifikasi tingkat risiko mengidap kanker paru-paru dengan nilai *accuracy* sebesar 97%, *precision* sebesar 97.26%, *recall* sebesar 97%, dan *f1-score* sebesar 96.98% berdasarkan hasil evaluasi metrik model yang terbentuk.

Kata kunci: Merokok, Kanker Paru-Paru, Kesehatan, Klasifikasi, *Random Forest*

Implementation Random Forest Algorithm to Classify Risk Level of Lung Cancer

Nehemia Gueldi

ABSTRACT

Lung cancer is one of the diseases that has become a major focus in global health due to its high mortality rate. Although the number of breast cancer cases in women surpasses lung cancer, lung cancer remains a leading cause of death. The main risk factors for lung cancer are smoking and exposure to air pollution. By constructing a number of decision trees separately and combining their results, the Random Forest algorithm can accurately classify and prevent overfitting. Testing was conducted using the Cancer Patient Datasets to measure the accuracy, precision, recall, and F1 score obtained. The results obtained in this study indicate that the Random Forest algorithm has been successfully implemented for classifying the risk level of developing lung cancer with an accuracy of 97%, precision of 97.26%, recall of 97%, and F1-score of 96.98% based on the evaluation metrics of the formed model.

Keywords: *Classification, Health, Lung Cancer, Random Forest, Smoking*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE

DAFTAR LAMPIRAN