



**KLASIFIKASI DATA MINING UNTUK MENDIAGNOSA PENYAKIT  
ISPA PADA ANAK DAN BALITA MENGGUNAKAN ALGORITMA  
NAÏVE BAYES DAN DECISION TREE**

**SKRIPSI**

**HOZANA AULIA**

**1910511100**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2023**



**KLASIFIKASI DATA MINING UNTUK MENDIAGNOSA PENYAKIT  
ISPA PADA ANAK DAN BALITA MENGGUNAKAN ALGORITMA  
NAÏVE BAYES DAN DECISION TREE**

**SKRIPSI**

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

**HOZANA AULIA**

**1910511100**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2023**

## PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Hozana Aulia

NIM : 1910511100

Tanggal : 26 Mei 2023

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya, maka saya bersedia di tuntutan dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 26 Mei 2023

Yang menyatakan,



(Hozana Aulia)

## **PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademis Universitas Pembangunan Nasional Veteran Jakarta,  
saya yang bertanda tangan di bawah ini:

Nama : Hozana Aulia  
NIM : 1910511100  
Fakultas : Ilmu Komputer  
Program Studi : S-1 Informatika

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

### **KLASIFIKASI DATA MINING UNTUK MENDIAGNOSA PENYAKIT ISPA PADA ANAK DAN BALITA MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN DECISION TREE**

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 mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat : Jakarta

Pada tanggal : 29 Mei 2023

Yang menyatakan



(Hozana Aulia)

## LEMBAR PENGESAHAN

Yang bertanda tangan di bawah ini:


Nama : Hozana Aulia  
NIM : 1910511100  
Program Studi : S-1 Informatika  
Judul Skripsi/TA : Klasifikasi Data Mining Untuk Mendiagnosa Penyakit  
ISPA Pada Anak Dan Balita Menggunakan Algoritma  
Naive Bayes Decision Tree

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




(Henki Bayu Seta, S.Kom., MTI.)

Penguji 1



(Theresia Wati, S.Kom., MTI.)  
Penguji 2



(Dr. Ermatita, M.Kom.)  
Pembimbing



(Dr. Ermatita, M.Kom.)  
Dekan



(Dr. Widya Cholil, M.T.)  
Ketua Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 26 Mei 2023



# KLASIFIKASI DATA MINING UNTUK MENDIAGNOSA PENYAKIT ISPA PADA ANAK DAN BALITA MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN DECISION TREE

Hozana Aulia

## ABSTRAK

Infeksi Saluran Pernapasan Akut (ISPA) masih menjadi penyakit saluran pernapasan yang paling banyak menular dengan kematian yang cukup tinggi di Dunia dengan kematian terbanyak dialami oleh anak dan balita. Di Indonesia sendiri ISPA menjadi salah satu penyakit yang tertinggi untuk penularannya. Penelitian dibidang kesehatan untuk mendiagnosa tingkat keparahan pasien penderita ISPA berdasarkan jenis penyakit ISPA yang diderita, gejala, usia jenis kelamin, dan masa idap sehingga perlu dilakukan untuk pengobatan lebih dini, guna untuk mencegah kematian pada anak dan balita akibat terlambatnya penanganan. Penelitian terkait prediksi penyakit menggunakan teknik *data mining* klasifikasi sudah secara luas digunakan. Untuk penelitian ini akan menggunakan algoritma Naïve Bayes dan Decision Tree serta akan dilakukan perbandingan untuk mendapatkan model terbaik diantara algoritma tersebut. Pengujian pertama menggunakan algoritma Naïve Bayes dengan tiga kali percobaan dengan membagi data latih dan data uji. Percobaan pertama dengan data latih 80% dan data uji 20% menghasilkan akurasi 92.30%, percobaan kedua dengan data latih 50% dan data uji 50% menghasilkan akurasi 88.88%, dan percobaan ketiga dengan data latih 90% dan data uji 10% menghasilkan akurasi 92.30%. Pengujian kedua menggunakan algoritma Decision Tree dengan tiga kali percobaan. Percobaan pertama dengan data latih 80% dan data uji 20% menghasilkan akurasi 88.46%, percobaan kedua dengan data latih 50% dan data uji 50% menghasilkan akurasi 90.47%, dan percobaan ketiga dengan data latih 90% dan data uji 10% menghasilkan akurasi 92.30%.

**Kata Kunci :** *ISPA, Data Mining, Klasifikasi, Naïve Bayes, Decision Tree*

# **CLASSIFICATION OF DATA MINING TO DIAGNOSE ARI DISEASE IN CHILDREN AND TODDLERS USING NAÏVE BAYES AND DECISION TREE**

**Hozana Aulia**

## **ABSTRACT**

Acute Respiratory Infection (ARI) is still the most widely transmitted respiratory disease with high mortality in the world, with most deaths occurring in children and toddlers. In Indonesia, ARI is one of the highest diseases for transmission. Research in the health sector is to diagnose the severity of patients with ISPA based on the type of ISPA disease they suffer, symptoms, age, sex, and duration of illness, so it is necessary to conduct earlier treatment, in order to prevent death in children and toddlers due to delays in treatment. Research related to disease prediction using classification data mining techniques has been widely used. For this research, the Naïve Bayes and Decision Tree algorithms will be used and comparisons will be made to get the best model among these algorithms. The first test uses the Naïve Bayes algorithm with three trials by dividing the training data and test data. The first experiment with 80% training data and 20% test data resulted in an accuracy of 92.30%, the second experiment with 50% training data and 50% test data resulted in an accuracy of 88.88%, and the third experiment with 90% training data and 10% test data resulted in an accuracy of 92.30%. The second test uses the Decision Tree algorithm with three trials. The first experiment with 80% training data and 20% test data produced an accuracy of 88.46%, the second experiment with 50% training data and 50% test data produced an accuracy of 90.47%, and the third experiment with 90% training data and 10% test data produced an accuracy of 92.30%.

**Keyword:** ARI, Data Mining, Classification, Naïve Bayes, Decision Tree

## KATA PENGANTAR

Puji syukur saya ucapkan kepada Allah SWT atas karunia dan ridha-Nya, sehingga Tugas Akhir dengan judul “Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Pada Anak dan Balita Menggunakan Algoritma Naïve Bayes dan Decision Tree” dapat dikerjakan tepat waktu. Penulis ingin mengucapkan terima kasih banyak kepada :

1. Orang tua serta keluarga yang terus mendoakan dan memberi semangat dan dukungan dalam berbagai hal .
2. Ibu Dr. Ermatita, M.Kom., selaku Dekan dan Dosen Pembimbing Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta yang telah banyak memberikan bimbingan serta saran dan masukan yang bermanfaat sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.
3. Ibu Widya Cholil, M.I.T selaku Kaprodi Informatika, Ibu Helena Nurramdhani Irmada, S.Pd., M.Kom selaku Dosen Pembimbing Akademik, Bapak Henki Bayu Seta, S.Kom., MTI dan Ibu Theresia Wati, S.Kom., MTI selaku Dosen Penguji serta seluruh Dosen dan Pegawai Fakultas Ilmu Komputer.
4. Orang terdekat penulis serta sahabat – sahabat penulis yang selalu ada saat dibutuhkan dan berkeluh kesah serta memberikan dukungan dan doa.
5. Teman – teman Informatika 2019 yang telah berjuang bersama dalam masa perkuliahan hingga dapat menyelesaikan Tugas Akhir ini.

Pada penelitian ini penulis masih banyak memiliki kekurangan dan masih jauh dari kata sempurna untuk penyusunan skripsi ini, sehingga penulis berharap mendapatkan saran dan kritik yang bersifat membangun untuk penulis. Akhir kata penulis ucapkan terima kasih.

Jakarta, 26 Mei 2023



(Hozana Aulia)



## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PENGESAHAN .....	v
KATA PENGANTAR .....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	2
1.5 Luaran Yang Diharapkan .....	3
1.6 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Data Mining.....	4
2.1.1 Definisi Data Mining.....	4
2.1.2 Tahapan Data Mining.....	4
2.1.3 Pengelompokan Data Mining.....	6
2.2 Infeksi Saluran Pernapasan Akut (ISPA) .....	7
2.3 Klasifikasi.....	8
2.4 Algoritma Naïve Bayes .....	8
2.5 Algoritma Decision Tree .....	10
2.6 Evaluasi .....	11
2.7 Python.....	13
2.8 Riset yang Relevan .....	13
BAB III METODOLOGI PENELITIAN .....	15
3.1 Alur Penelitian.....	15
3.2 Studi Literatur.....	15
3.3 Identifikasi Masalah .....	16
3.4 Pengumpulan Data .....	16

3.5	Pra-proses Data.....	16
3.5.1	Data Cleaning.....	16
3.5.2	Transformasi Data.....	17
3.6	Pembagian Data.....	18
3.7	Klasifikasi.....	18
3.8	Evaluasi .....	18
3.9	Perbandingan .....	18
3.10	Alat Bantu Penelitian.....	19
3.11	Jadwal Penelitian .....	19
BAB IV HASIL DAN PEMBAHASAN .....		20
4.1	Pengumpulan Data .....	20
4.2	Pra-proses Data.....	20
4.3	Cleaning Data .....	20
4.4	Transformasi Data .....	21
4.5	Pembagian Data.....	22
4.6	Klasifikasi.....	22
4.6.1	Naïve Bayes .....	22
4.6.2	Decision Tree .....	35
4.7	Perbandingan .....	52
BAB V_ PENUTUP.....		55
5.1	Kesimpulan.....	55
5.2	Saran .....	55
DAFTAR PUSTAKA .....		56
LAMPIRAN.....		59
Lampiran 1. Surat Permohonan Izin Riset .....		59
Lampiran 2. Surat Keterangan RSUD .....		60
Lampiran 3. Wawancara Pra-Penelitian.....		61
Lampiran 4. Code Naïve Bayes.....		62
Lampiran 5. Code Decision Tree.....		64
Lampiran 6. Hasil Turnitin .....		67
Lampiran 7. Biodata Mahasiswa .....		78

## DAFTAR GAMBAR

Gambar 2. 1. Data Mining sebagai Proses dalam KDD.....	5
Gambar 3. 1. Alur Penelitian.....	15
Gambar 4. 1. Pohon Keputusan.....	44
Gambar 4. 2. Pohon Keputusan Node Kiri .....	47
Gambar 4. 3. Pohon Keputusan Node Kanan .....	49

## DAFTAR TABEL

Tabel 2. 1. Confusion Matrix .....	11
Tabel 3. 1. Faktor Penentu Tingkatan ISPA .....	17
Tabel 3. 2. Jadwal Penelitian.....	19
Tabel 4. 1. Dataset ISPA .....	20
Tabel 4. 2. Dataset ISPA Setelah Transformasi .....	21
Tabel 4. 3. Dataset Acak .....	23
Tabel 4. 4. Probabilitas Atribut Jenis Kelamin .....	24
Tabel 4. 5. Probabilitas atribut usia.....	24
Tabel 4. 6. probabilitas atribut Diagnosa .....	24
Tabel 4. 7. probabilitas atribut Lama Idap .....	24
Tabel 4. 8. probabilitas Gejala .....	24
Tabel 4. 9. Probabilitas atribut jenis kelamin.....	25
Tabel 4. 10. Probabilitas atribut usia.....	25
Tabel 4. 11. probabilitas atribut Diagnosa .....	25
Tabel 4. 12. probabilitas atribut Lama Idap .....	26
Tabel 4. 13. probabilitas atribut Gejala.....	26
Tabel 4. 14. Confusion Matrik Dengan Data Sampel .....	32
Tabel 4. 15. Percobaan 20% data testing .....	33
Tabel 4. 16. Percobaan dengan 10% data testing.....	34
Tabel 4. 17. percobaan 50% data testing .....	35
Tabel 4. 18. Sampel Data .....	36
Tabel 4. 19. Perhitungan Nilai Gain dan Entropy .....	37
Tabel 4. 20. Data Berdasarkan Lama Idap dengan Nilai 2 .....	44
Tabel 4. 21. Perhitungan Gain dan Entropy .....	44
Tabel 4. 23 Data Berdasarkan Gejala dengan Nilai 2 .....	48
Tabel 4. 24 Perhitungan Nilai Gain dan Entropy .....	48
Tabel 4. 24. Percobaan 20% data testing .....	50
Tabel 4. 25. Percobaan 50% data testing .....	51
Tabel 4. 26. Percobaan 10% data testing .....	51
Tabel 4. 27. Perbandingan Performa Klasifikasi .....	52