

**SISTEM DIAGNOSIS PENYAKIT COVID-19  
MENGUNAKAN ALGORITMA C5.0 DENGAN  
MENERAPKAN METODE *REDUCE ERROR PRUNING (REP)***

**RINGKASAN SKRIPSI**



Oleh :

**MAULANA IHSANUR ROZAO**

**NPM. 1634010047**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2022**

**LEMBAR PENGESAHAN  
SKRIPSI**

**Judul : SISTEM DIAGNOSIS PENYAKIT COVID-19 MENGGUNAKAN  
ALGORITMA C5.0 DENGAN MENERAPKAN METODE  
REDUCE ERROR PRUNING (REP)**

**Oleh : MAULANA IHSANUR ROZAQ**

**NPM : 1634010047**

**Telah Diseminarkan Dalam Ujian Skripsi Pada :**

**Hari Rabu, Tanggal 7 Desember 2022**

**Mengetahui**

**Dosen Pembimbing**

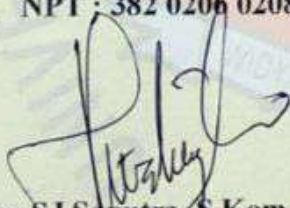
1.



**Fetty Tri Anggraeni, S.Kom., M.Kom**

**NPT : 382 0206 0208 1**

2.



**Wahyu SJ Saputra, S.Kom., M.Kom**

**NIPPPK : 19860825 2021211 003**

**Dosen Penguji**

1.



**Yisti Vita Via, S.ST. M.Kom.**

**NIPPPK : 19860425 2021212 001**

2.



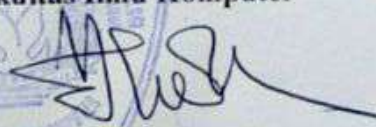
**Eka Prakarsa Mandvartha, S.T., M.Kom**

**NIP : 19880525 2018031 001**

**Menyetujui**

**Dekan**

**Fakultas Ilmu Komputer**

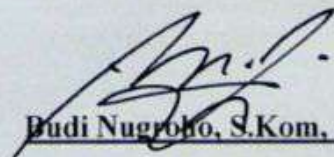


**Dr. Ir. Ni-Ketut Sari, MT**

**NPT : 19650731 1199203 2 001**

**Koordinator Program Studi**

**Teknik Informatika**



**Budi Nugroho, S.Kom, M.Kom**

**NPT : 3 8009 05 0205 1**

## SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Informatika UPN "Veteran" Jawa Timur, yang bertandatangan di bawah ini:

Nama : MAULANA IHSANUR ROZAQ

NPM : 1634010047

Menyatakan bahwa Judul Skripsi/ Tugas Akhir yang saya ajukan dan akan dikerjakan, yang berjudul:

**"SISTEM DIAGNOSIS PENYAKIT COVID-19 MENGGUNAKAN ALGORITMA C5.0 DENGAN MENERAPKAN METODE REDUCE ERROR PRUNING (REP)"**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk dan atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 07 Desember 2022

Hormat Saya,



**MAULANA IHSANUR ROZAQ**

**NPM. 1634010047**

# **SISTEM DIAGNOSIS PENYAKIT COVID-19 MENGUNAKAN ALGORITMA C5.0 DENGAN MENERAPKAN METODE *REDUCE ERROR PRUNING (REP)***

**Nama Mahasiswa : Maulana Ihsanur Rozaq**  
**NPM : 1634010047**  
**Program Studi : Teknik Informatika**  
**Dosen Pembimbing : Fetty Tri Anggraeny, S.Kom., M.Kom**  
**Wahyu SJ Saputra, S.Kom., M.Kom**

## **Abstrak**

Coronavirus adalah suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Infeksi COVID-19 dapat menimbulkan gejala ringan, sedang atau berat. Gejala klinis utama yang muncul yaitu demam (suhu > 38C), batuk dan kesulitan bernapas. Dalam keadaan yang seperti ini, adanya suatu sistem yang dapat mengetahui kondisi yang dapat mengidentifikasi seseorang positif atau negatif covid-19 pastinya akan sangat membantu. Namun, dibutuhkan suatu metode yang memiliki tingkat akurasi yang tinggi dalam penerapannya kedalam sistem. Oleh karena itu, digunakan algoritma C5.0 untuk dapat melakukan diagnosis covid-19 berdasarkan dari gejala yang dialami seseorang.

Algoritma C5.0 adalah salah satu algoritma klasifikasi, penyempurnaan dari algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Pada algoritma C5.0 terdapat suatu metode yang disebut dengan *boosting*, metode inilah yang menjadi salah satu kelebihan dari C5.0. Dengan adanya *boosting*, mampu meningkatkan akurasi dari pohon keputusan yang dibentuk oleh algoritma C5.0, dimana cara kerjanya yaitu dengan mengacak *dataset* hingga menemukan akurasi yang lebih tinggi.

Setelah terbentuknya pohon keputusan yang terbaik, akan dilakukan percobaan sebanyak sepuluh kali percobaan untuk mendapatkan nilai akurasi yang terbaik. Dari sepuluh percobaan tersebut diambil pohon keputusan yang memiliki nilai akurasi paling besar. Pohon keputusan tersebut kemudian diuji dengan 545 data tes dan menghasilkan nilai *accuracy* sebesar 90,46 %, *recall* sebesar 94,76 % dan *precision* sebesar 93,48%.

***Kata kunci:*** covid-19, C5.0, *boosting*, pohon keputusan, coronavirus

## **KATA PENGANTAR**

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena dengan karunia-Nya, penulis dapat menyelesaikan skripsi ini yang berjudul “Sistem Diagnosis Penyakit Covid-19 Menggunakan Algoritma *C5.0* Dengan Menggunakan Metode *Reduce Error Pruning (REP)*”. Tujuan skripsi ini dibuat sebagai salah satu persyaratan untuk mendapatkan gelar Sarjana Komputer di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis berharap dengan penyusunan skripsi ini dapat memberikan manfaat bagi teman-teman maupun pembaca pada umumnya. Penulis menyadari dalam penulisan skripsi ini masih jauh dari kata sempurna, dan banyak kekurangan baik dalam pembahasan materi maupun dalam penyusunan laporan. Sehingga penulis sangat mengharapkan kritik dan saran yang bersifat membangun yang nantinya dapat dijadikan pelajaran oleh penulis maupun pembaca dikemudian hari.

Surabaya, 26 September 2022

Penulis

## DAFTAR ISI

ABSTRAK .....	ii
KATA PENGANTAR .....	iii
UCAPAN TERIMAKASIH .....	iv
DAFTAR ISI .....	vi
DAFTAR TABEL .....	ix
DAFTAR GAMBAR .....	x
DAFTAR KODE .....	xi
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan .....	3
1.4. Manfaat .....	3
1.5. Batasan Masalah .....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1. Penelitian Pendukung .....	4
2.2. <i>Decision Tree</i> .....	5
2.3. Algoritma C5.0 .....	6
2.4. <i>Reduce Error Pruning</i> (REP) .....	13
2.5. <i>Boosting</i> .....	15
2.6. <i>Confusion Matrix</i> .....	17
BAB III METODOLOGI .....	19
3.1. Langkah-langkah Penelitian .....	19
3.2. Dataset Penelitian .....	20
3.3. <i>Preprocessing Data</i> .....	21

3.4.	Pembagian Data .....	22
3.5.	Algoritma C5.0 .....	23
3.5.1.	<i>Pruning</i> .....	31
3.5.2.	<i>Boosting</i> .....	33
3.6.	Skenario Pengujian Sistem .....	35
BAB IV HASIL DAN PEMBAHASAN .....		36
4.1.	Implementasi Data .....	36
4.2.	Implementasi Desain <i>Interface</i> .....	37
4.2.1.	Tampilan Awal (Beranda) .....	37
4.2.2.	Tampilan Menu <i>Dataset</i> .....	38
4.2.3.	Tampilan Menu Data Latih .....	38
4.2.4.	Tampilan Menu Data Tes .....	39
4.2.5.	Tampilan Menu Tabel Perhitungan C5.0 .....	39
4.2.6.	Tampilan Menu Tabel Perhitungan <i>Pruning</i> .....	40
4.2.7.	Tampilan Menu Tabel <i>Boosting</i> .....	42
4.2.8.	Tampilan Menu Keputusan Akhir .....	44
4.2.9.	Tampilan Menu Cek .....	48
4.3.	Pembahasan .....	49
4.3.1.	Perhitungan <i>Entropy</i> .....	49
4.3.2.	Perhitungan <i>Information Gain</i> .....	50
4.3.3.	Perhitungan <i>Gain Ratio</i> .....	51
4.3.4.	Perhitungan <i>Pruning</i> .....	51
4.3.5.	Perhitungan <i>Boosting</i> .....	53
4.4.	Pengujian Sistem .....	54
BAB V KESIMPULAN DAN SARAN .....		58
5.1.	Kesimpulan .....	58

5.2. Saran.....	59
DAFTAR PUSTAKA.....	60



## DAFTAR TABEL

<b>Tabel 2.1</b> Contoh Data Pendapatan Masyarakat .....	7
<b>Tabel 2.2</b> Pengelompokan Data .....	8
<b>Tabel 2.3</b> Nilai <i>Entropy</i> Seluruh Atribut .....	9
<b>Tabel 2.4</b> Hasil <i>Entropy</i> dan <i>Information Gain</i> .....	11
<b>Tabel 2.5</b> Hasil <i>Gain Ratio</i> .....	12
<b>Tabel 2.6</b> Tabel <i>Confusion Matrix</i> .....	18
<b>Tabel 3.1</b> Contoh Data .....	25
<b>Tabel 3.2</b> Klasifikasi data latih .....	26
<b>Tabel 3.3</b> Nilai <i>entropy</i> , <i>information gain</i> dan <i>gain ratio</i> . .....	28
<b>Tabel 4.1</b> Tabel Akurasi Hasil Percobaan .....	55
<b>Tabel 4.2</b> Tabel <i>Confussion Matrix</i> Data Tes .....	57

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Pohon Keputusan Node Pertama .....	13
<b>Gambar 3.1</b> Langkah-langkah Penelitian .....	19
<b>Gambar 3.2</b> Langkah-langkah Algoritma C5.0.....	24
<b>Gambar 3.3</b> <i>Parent Node</i> dengan Atribut <i>Abroad Travel</i> .....	30
<b>Gambar 3.4</b> <i>Flowchart</i> Proses <i>Pruning</i> .....	31
<b>Gambar 3.5</b> Hasil Proses <i>Pruning</i> .....	33
<b>Gambar 3.6</b> <i>Flowchart</i> proses <i>Boosting</i> .....	34
<b>Gambar 4.1</b> Tampilan Beranda .....	37
<b>Gambar 4.2</b> Tampilan Menu <i>Dataset</i> .....	38
<b>Gambar 4.3</b> Tampilan Menu Data Latih.....	39
<b>Gambar 4.4</b> Tampilan Menu Data Tes .....	39
<b>Gambar 4.5</b> Tampilan Menu Hitung .....	40
<b>Gambar 4.6</b> Tampilan Menu <i>Pruning</i> Tabel Perhitungan <i>Pruning</i> .....	41
<b>Gambar 4.7</b> Tampilan Menu <i>Pruning</i> Tabel Data <i>Pruning</i> .....	42
<b>Gambar 4.8</b> Tampilan Menu <i>Boosting</i> .....	43
<b>Gambar 4.9</b> Tampilan Hasil Uji Data Tes .....	43
<b>Gambar 4.10</b> Tampilan Konfirmasi <i>Update</i> Data Acuan .....	44
<b>Gambar 4.11</b> Tampilan Menu Keputusan Akhir Hasil Akurasi .....	44
<b>Gambar 4.12</b> Grafik Hasil <i>Boosting</i> .....	45
<b>Gambar 4.13</b> Tampilan Pohon Keputusan.....	46
<b>Gambar 4.14</b> Tampilan Hasil Uji Menggunakan Data Tes .....	47
<b>Gambar 4.15</b> Tampilan Konfirmasi <i>Update</i> Data Acuan.....	47
<b>Gambar 4.16</b> Tampilan <i>Form Input</i> Gejala .....	48
<b>Gambar 4.17</b> Tampilan Hasil Pengecekan Gejala Baru .....	49
<b>Gambar 4.18</b> Grafik Hasil Percobaan.....	54
<b>Gambar 4.19</b> Hasil Uji Data Tes.....	56

## DAFTAR KODE

<b>Kode Program 4.1</b> Pengelompokkan Data Latih .....	50
<b>Kode Program 4.2</b> Perhitungan <i>Entropy</i> Kelas .....	50
<b>Kode Program 4.3</b> Perhitungan <i>Entropy</i> Seluruh Atribut .....	50
<b>Kode Program 4.4</b> Perhitungan <i>Information Gain</i> .....	51
<b>Kode Program 4.5</b> Perhitungan <i>Gain Ratio</i> .....	51
<b>Kode Program 4.6</b> Fungsi Perhitungan <i>Pruning</i> .....	52
<b>Kode Program 4.7</b> Pengacakan Data .....	53
<b>Kode Program 4.8</b> Fungsi Perhitungan <i>Boosting</i> .....	54