

**ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT  
HEPATITIS C MENGGUNAKAN ALGORITMA DECISION TREE  
C.45 DENGAN PARTICLE SWARM OPTIMIZATION**

**TESIS**



**Disusun Oleh  
Fely Dany Prasetya  
2021211014**

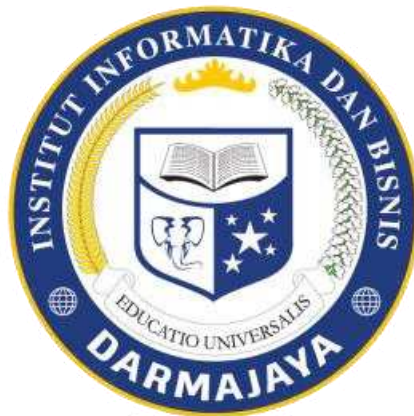
**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA  
BANDAR LAMPUNG**

**2022**

**ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT  
HEPATITIS C MENGGUNAKAN ALGORITMA DECISION TREE  
C.45 DENGAN PARTICLE SWARM OPTIMIZATION**

**TESIS**

**Diajukan sebagai Salah Satu Syarat untuk Mencapai Gelar  
Magister Teknik Informatika  
Pada Program Studi Teknik Informatika**



**Disusun Oleh  
Fely Dany Prasetya  
2021211014**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA  
BANDAR LAMPUNG  
2022**



## PERNYATAAN

Saya yang bertanda tangan dibawah ini, menyatakan bahwa tesis yang saya ajukan ini adalah hasil karya saya sendiri, tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi atau karya yang pernah ditulis atau diterbitkan orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Karya ini adalah milik saya dan pertanggungjawaban sepenuhnya berada di pundak saya.

Bandar Lampung, Desember 2022


Fely Dany Prasetya

NPM. 202121014

## HALAMAN PERSETUJUAN

Judul Tesis : "ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT  
HEPATITIS C MENGGUNAKAN ALGORITMA DECISION  
TREE C.45 DENGAN *PARTICLE SWARM OPTIMIZATION*"

Nama Mahasiswa : Fely Dany Prasetya

NPM : 202121014

Program Studi : Magister Teknik Informatika



Menyetujui,

Dosen Pembimbing

A handwritten signature in black ink, appearing to read "Handoyo".

Dr. Handoyo Widi Nugroho, S.Kom., M.T.I  
NIK. 004005002

A handwritten signature in black ink, appearing to read "Joko Triloka".

Joko Triloka, S.Kom., M.T., Ph.D  
NIK. 00610303

## HALAMAN PENGESAHAN

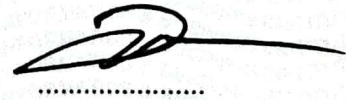
Telah diuji dan dipertahankan didepan Tim Penguji Tugas Akhir Program Studi  
Magister Teknik Informatika IIB Darmajaya dan dinyatakan diterima untuk  
memenuhi syarat guna memperoleh gelar Magister

Mengesahkan,

1. Tim Penguji

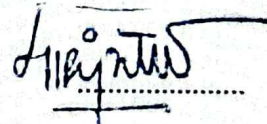
Tanda Tangan

Anggota 1 : Joko Triloka, S.Kom., M.T., Ph.D



.....

Anggota 2 : Priyanto, S.Kom., M.M., Ph.D



.....

Dekan Fakultas Ilmu Komputer



Dr. Sutedi, S.Kom., M.T.I

NIK/00600303

Tanggal Lulus Ujian Tesis:

## HALAMAN PERSEMBAHAN

Dengan mengucapkan Alhamdulillahirabbil‘alamin Karya ini penulis persembahkan kepada :

Kedua orang tua yang telah membesarkan dan mendidiku dengan penuh cinta serta kasih sayang tiada hentinya memberikan do’a yang selalu mengiringi langkahku dan tak pernah bosan memberiku semangat, bimbingan dan nasehat di setiap langkah kehidupanku

Untuk

Istriku Tercinta, Adik-adikku, Saudara-Saudari, serta Teman-teman almamaterku yang telah memberikan doa dan semangat baik itu moril maupun materil.

## **MOTTO**

“Bersyukur & Selalu Berbuat Baik Kepada Orang Lain”

-(FELY DANY PRASETYA)-

## DAFTAR RIWAYAT HIDUP

### DATA PRIBADI

1. Nama : Fely Dany Prasetya
2. Tempat ,Tanggal Lahir : Panutan, 29 Maret 1990
3. Alamat : Gumukmas RT 002/ RW 004  
Kecamatan Pagelaran Kabupaten Pringsewu
4. Domisili : Pringsewu
5. Jenis Kelamin : Laki-laki
6. Agama : Islam
7. Status : Nikah
8. Kewarganegaraan : Indonesia
9. Suku : Jawa
10. Telepon : 082182549494
11. Email : [felydanyprasetya@gmail.com](mailto:felydanyprasetya@gmail.com)

### RIWAYAT PENDIDIKAN

1. (2002) Lulus SD N 4 Patoman
2. (2005) Lulus SMP N 1 Pagelaran
3. (2008) Lulus SMK YPT Pringsewu
4. (2013) Lulus S1 STMIK Teknokrat

Bandar Lampung, Desember 2022



**Fely Dany Prasetya**

NPM. 202121014



## ABSTRAK

# ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT HEPATITIS C MENGUNAKAN ALGORITMA DECISION TREE C.45 DENGAN PARTICLE SWARM OPTIMIZATION

Oleh

Fely Dany Prasetya

Hepatitis sendiri merupakan penyakit peradangan pada hati (liver) yang dapat disebabkan oleh faktor genetik, infeksi virus, alkohol, dan obat-obatan. Berdasarkan laporan global oleh Organisasi. Secara global, diperkirakan 58 juta orang memiliki infeksi virus hepatitis C kronis, dengan sekitar 1,5 juta infeksi baru terjadi setiap tahun. Diperkirakan ada 3,2 juta remaja dan anak-anak dengan infeksi hepatitis C kronis. Pada tahun 2019 WHO memperkirakan sekitar 290.000 orang meninggal karena hepatitis C, sebagian besar karena sirosis dan karsinoma hepatoseluler (kanker hati primer). Dalam penelitian ini metode yang digunakan adalah teknik klasifikasi dalam data mining menggunakan algoritma Decision Tree C. 45 yang dioptimasi dengan Particle Swarm Optimization (PSO). Tahapan penelitian ini menerapkan CRISP-DM dimulai dengan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan diakhiri dengan evaluasi. Hasil yang diperoleh dari pengujian data hepatitis dengan kategori \*0=Donor Darah, 0s=dugaan Donor Darah, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis) pada penggunaan metode C4.5 Decision Tree Algorithm hanya pada klasifikasi untuk mendapatkan nilai akurasi yang tinggi. yang paling optimal adalah 87,15%. Kemudian pengujian penggunaan algoritma Decision Tree C4.5 dengan optimasi Algoritma PSO menghasilkan nilai akurasi paling optimal sebesar 87,48% dengan demikian optimasi dengan Algoritma PSO dapat meningkatkan akurasi algoritma DecisionTree C4.5 sehingga menghasilkan hasil yang lebih akurasi optimal dengan selisih akurasi 0,33%.

Kata Kunci: Data Mining,;Algoritma Decision Tree C.45,;PSO,;Hepatitis C.

## ABSTRACT

### DATA MINING ANALYSIS FOR HEPATITIS C PREDICTION USING DECISION TREE C.45 ALGORITHM WITH PARTICLE SWARM OPTIMIZATION

By

Fely Dany Prasetya

Hepatitis itself is an inflammatory disease of the liver (liver) that can be caused by genetic factors, viral infections, alcohol, and drugs. Based on global reports by the Organization. Globally, an estimated 58 million people have chronic hepatitis C virus infection, with approximately 1.5 million new infections occurring each year. It is estimated that there are 3.2 million adolescents and children with chronic hepatitis C infection. In 2019 WHO estimated that around 290,000 people died from hepatitis C, mostly from cirrhosis and hepatocellular carcinoma (primary liver cancer). In this study, the method used is a classification technique in data mining using the Decision Tree C. 45 algorithm which is optimized by Particle Swarm Optimization (PSO). This research stage applies CRISP-DM starting with business understanding, data understanding, data preparation, modeling, and ending with evaluation. The results obtained from testing the hepatitis data with the category \*0=Blood Donor, 0s=suspected Blood Donor, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis) on the use of the C4.5 Decision Tree Algorithm method only on classification to obtain accurate values. tall. the most optimal is 87.15%. Then testing the use of the Decision Tree C4.5 algorithm with the optimization of the PSO Algorithm produces the most optimal accuracy value of 87.48% thus optimization with the PSO Algorithm can increase the accuracy of the DecisionTree C4.5 algorithm so as to produce more optimal accuracy results with an accuracy difference of 0.33%.

Keywords : Data Mining,;Decision Tree C.45 Algorithm,;PSO,;Hepatitis C.

## PRAKATA

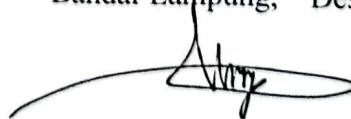
Puji dan syukur kepada tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga saya telah menyelesaikan tesis yang berjudul “**ANALISA DATA MINING UNTUK PREDIKSI PENYAKIT HEPATITIS C MENGGUNAKAN ALGORITMA DECISION TREE C.45 DENGAN *PARTICLE SWARM OPTIMIZATION***”. Tesis ini disusun sebagai syarat untuk menyelesaikan jenjang studi Strata Dua (S2) Teknik Informatika, Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung. Penyusun banyak mendapatkan pengarahan, bimbingan dan bantuan sehingga penyusun dapat menyelesaikan tesis ini. Dalam kesempatan ini, perkenankan penyusun menyampaikan rasa terimakasih kepada :

1. Bapak Dr. Ir. Firmansyah YA, M.B.A., M.Sc selaku Rektor IIB Darmajaya
2. Bapak Dr. Sutedi, S.Kom, M.T.I Selaku Dekan Fakultas Ilmu Komputer IIB Darmajaya
3. Bapak Dr. Handoyo Widi Nugroho, S.Kom., M.T.I selaku pembimbing yang telah membimbing dengan sabar dan mengarahkan penulis selama penyusunan tesis ini dari awal hingga selesai.
4. Bapak Joko Triloka, S.Kom., MT., Ph.D selaku ketua program studi Magister Teknik Informatika IIB Darmajaya.
5. Para Dosen dan karyawan Pascasarjana IIB Darmajaya yang telah membimbing saya dengan ilmu-ilmu yang bermanfaat dan memotivasi dalam penyelesaian tesis ini.

Penyusun sadar akan kekurangan yang mungkin masih ada dalam penulisan tesis ini maka kesediaan memberikan kritik dan saran yang membangun untuk menciptakan karya yang lebih baik lagi dimasa yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis maupun pembaca.

Bandar Lampung, Desember 2022



**Fely Dany Prasetya**

NPM. 202121014

## DAFTAR ISI

Halaman

|   |     |
|---|-----|
| HALAMAN JUDUL .....                         | i   |
| HALAMAN PERSETUJUAN .....                   | iv  |
| HALAMAN PENGESAHAN .....                    | v   |
| DAFTAR ISI.....                             | xii |
| DAFTAR TABEL .....                          | xiv |
| DAFTAR GAMBAR .....                         | xv  |
| BAB I PENDAHULUAN                           |     |
| 1.1 Latar Belakang .....                    | 1   |
| 1.2 Batasan Masalah .....                   | 2   |
| 1.3 Rumusan Masalah.....                    | 2   |
| 1.4 Tujuan Penelitian .....                 | 2   |
| 1.5 Manfaat Penelitian .....                | 3   |
| 1.6 Sistematika Penulisan .....             | 3   |
| BAB II TINJAUAN PUSTAKA                     |     |
| 2.1 Penelitian Terkait .....                | 4   |
| 2.2 Hepatitis.....                          | 5   |
| 2.3 Data Mining.....                        | 6   |
| 2.4 Klasifikasi.....                        | 7   |
| 2.5 Decision Tree C.45 .....                | 8   |
| 2.6 Particle Swarm Optimization (PSO) ..... | 9   |
| 2.7 Confusion Matrix.....                   | 9   |
| 2.8 Feature Selection .....                 | 11  |
| 2.9 Metode CRISP–DM .....                   | 12  |
| 3.0 RapidMiner.....                         | 15  |

## BAB III METODE PENELITIAN

|                              |    |
|------------------------------|----|
| 3.1 Metode Penelitian .....  | 18 |
| 3.2 Kerangka Pemikiran ..... | 21 |
| 3.4 Desain Penelitian .....  | 22 |
| 3.3 Tahapan Penelitian ..... | 23 |
| 3.5 Alat dan Bahan .....     | 24 |

## BAB IV HASIL DAN PEMBAHASAN

|   |    |
|---|----|
| 4.1 Persiapan Data .....  | 24 |
| 4.2 Pemodelan .....   | 26 |
| 4.2.1 Penelitian Menggunakan Algoritma Decision Tree C4.5 .....                 | 26 |
| 4.2.1 Penelitian Menggunakan Algoritma Decision Tree C4.5+PSO .....             | 31 |
| 4.3 Pembandingan Pengujian Model Algoritma C4.5 dengan Algoritma C4.5+PSO ..... | 34 |
| 4.4 Evaluasi .....  | 35 |

## BAB V KESIMPULAN DAN SARAN

|                      |    |
|----------------------|----|
| 5.1 Kesimpulan ..... | 36 |
| 5.2 Saran .....      | 36 |

## DAFTAR PUSTAKA

## LAMPIRAN

## DAFTAR TABEL

| Tabel   | Halaman |
|---|---------|
| 3.1 Atribut-atribut Dataset .....                             | 19      |
| 4.1 Atribut Dataset .....                                     | 25      |
| 4.2 Hasil Akurasi Decision Tree C4.5.....                     | 28      |
| 4.3 Nilai Entrophy Total Pada Dataset .....                   | 29      |
| 4.4 Nilai Entrophy Dan Gain Pada Atribut Age Dan Sex .....    | 29      |
| 4.5 Hasil Akurasi Decision Tree C4.5+ PSO.....                | 33      |
| 4.6 Tabel Atribut yang digunakan sesudah PSO .....            | 33      |
| 4.7 Perbandingan Pengujian Sebelum dan Sesudah Optimasi ..... | 34      |

## DAFTAR GAMBAR

| Gambar  | Halaman |
|---|---------|
| 2.1 Metrik Kinerja Model.....                                   | 10      |
| 2.2 Tahapan Metode CRISP-DM .....                               | 12      |
| 2.3 Tools Rapidminer .....                                      | 16      |
| 3.1 Tahapan Data Mining .....                                   | 18      |
| 3.2 Tahapan Pemodelan .....                                     | 20      |
| 3.3 Kerangka Pemikiran.....                                     | 21      |
| 3.4 Skema Riset.....  | 22      |
| 3.5 Tahapan Penelitian .....                                    | 22      |
| 4.1 Dataset Hepatitis C .....                                   | 24      |
| 4.2 Model Rapidminer Tanpa Menggunakan Seleksi Fitur.....       | 26      |
| 4.3 Confusion Matrix Tanpa Menggunakan Seleksi Fitur.....       | 26      |
| 4.4 Tampilan Pohon Keputusan Decision Tree C4.5 .....           | 28      |
| 4.5 Tampilan Deskripsi Pohon Keputusan Decision Tree C4.5 ..... | 28      |
| 4.6 Model Rapidminer Menggunakan Seleksi Fitur.....             | 31      |
| 4.7 Confusion Matrix Menggunakan Seleksi Fitur .....            | 31      |
| 4.8 Atribusi Weight Hasil Akurasi C.45 + PSO.....               | 35      |
| 4.9 Weight Visual Hasil Akurasi Decision Tree C.45 + PSO .....  | 35      |