Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022
Institut Informatika dan Bisnis Darmajaya, 27 Agustus 2022
ISSN: 2598-0256, E-ISSN: 2598-0238

Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree *C.45* Dengan*Particle Swarm Optimization*

Fely Dany Prasetya[1a, *], Handoyo Widi Nugroho[2b], Joko Triloka [2c]

[a] Program Studi Magister Teknik Informatika, Fakultas Ilmu Komputer IIB Darmajaya
[a] fely.prasetya.2021211014@mail.darmajaya.ac.id
[b]handoyo.wn@darmajaya.ac.id
[c]joko.triloka@darmajaya.ac.id

Abstract

Hepatitis itself is an inflammatory disease of the liver (liver) that can be caused by genetic factors, viral infections, alcohol and drugs. Based on worldwide reports from the organization. An estimated 58 million people worldwide are chronically infected with the hepatitis C virus, with approximately 1.5 million new infections each year. It is estimated that there are 3.2 million adolescents and children with chronic hepatitis C infection. According to WHO estimates, in 2019 about 290,000 people died from hepatitis C, mainly from cirrhosis and hepatocellular carcinoma (primary liver cancer). The method used in this study is a classification technique in data mining using the Particle Swarm Optimization (PSO) optimized Decision Tree C.45 algorithm. The results of testing hepatitis data with the category *0=blood donor, 0s=blood donor suspect, 1=hepatitis, 2=fibrosis, 3=cirrhosis) applying the C4.5 decision tree algorithm procedure obtained accurate values only when categorized by. high. the optimum is 99.35%. Testing the use of the DecisionTree C4.5 algorithm with the PSO algorithm optimization gives the most optimal accuracy value of 99.67%, so the optimization with the PSO algorithm can increase the accuracy of the DecisionTree C4.5 increase algorithm by one achieve better performance Accuracy results with an accuracy difference of 0.32%.

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

Abstrak

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). 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 99,35%. Kemudian pengujian penggunaan algoritma Decision Tree C4.5 dengan optimasi Algoritma PSO menghasilkan nilai akurasi paling optimal sebesar 99,67% dengan demikian optimasi dengan Algoritma PSO dapat meningkatkan akurasi algoritma DecisionTree C4.5 sehingga menghasilkan hasil yang lebih akurasi optimal dengan selisih akurasi 0,32%.

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

1. PENDAHULUAN

Virus hepatitis C (HCV) adalah virus RNA tunggal yang secara resmi diidentifikasi pada April 1989 sebagai penyebab utama hepatitis non-A non-B.(Lanini et al., 2016). Hepatitis sendiri merupakan penyakit peradangan pada

ISSN: 2598-0256, E-ISSN: 2598-0238

hati (liver) yang dapat disebabkan oleh faktor genetik, infeksi virus, alkohol dan obat-obatan. Menurut laporan global dari Organisasi Kesehatan Dunia (WHO), diperkirakan 58 juta orang di seluruh dunia menderita infeksi hepatitis C kronis, dengan sekitar 1,5 juta infeksi baru 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, terutama karena sirosis dan karsinoma hepatoseluler (kanker hati primer).(WHO, 2022). Virus hepatitis C (VHC) merupakan salah satu virus penyebab hepatitis dan dianggap sebagai virus penyebab hepatitis yang paling mematikan. Kebanyakan orang yang terinfeksi virus hepatitis C tidak menunjukkan gejala. Banyak orang tidak menyadari bahwa mereka telah tertular virus hepatitis C sampai hati mereka rusak parah.(Alhawaris, 2019).

Dengan pesatnya perkembangan teknologi, penggunaan sistem informasi yang terkomputerisasi semakin meluas di berbagai bidang, termasuk bidang medis dan kesehatan. Sektor kesehatan telah mampu menghasilkan sejumlah besar data dan jumlah ini akan terus bertambah. Jumlah data yang meningkat ini memerlukan metode otomatis untuk mengekstrak data ini jika perlu (Milovic & Milovic, 2012). Jumlah data pasien dapat diolah dengan menggunakan teknik data mining. Data mining adalah solusi yang memungkinkan kami menemukan konten informasi tersembunyi dalam bentuk pola dan aturan dari kumpulan data besar dengan cara yang dapat dipahami (Latu Handarko, 2015). Penggunaan teknologi ini dapat diterapkan dalam memprediksi pasien yang terinfeksi hepatitis C untuk mengidentifikasi pasien secara cepat pada tahap awal. Deteksi Pasien Hepatitis C Menggunakan Anti-HCV Anti-HCV adalah salah satu tes yang dilakukan untuk memeriksa antibodi HCV dalam serum pasien. Antibodi ini terbentuk dalam serum saat pasien terinfeksi virus hepatitis C. Deteksi dini bisa dilakukan dengan rapid test dan hasilnya terlihat setelah 15 menit. Dalam ilmu komputer, data mining merupakan ilmu yang dapat membantu memprediksi pasien hepatitis C. Clinical data mining adalah penerapan teknik data mining untuk mengungkap data medis dan klinis. Dengan metode ini, kondisi masa depan pasien dapat diprediksi berdasarkan data pasien lain dan data observasi masa lalu. Salah satu metode prediksi adalah klasifikasi. Kami menguji beberapa metode klasifikasi untuk mengkonfirmasi keakuratan hasil prediksi hepatitis.(Studi & Informatika, n.d.).

Di bidang ilmu komputer, beberapa penelitian telah dilakukan untuk memprediksi penyakit hepatitis C dengan teknik data mining menggunakan studi algoritma Decision Tree C.45 berjudul Menerapkan Teknik Data Mining untuk Mengklasifikasikan Pasien Terduga Infeksi Virus Hepatitis C oleh algoritma Safdari et al. yaitu SVM, Nave Bayes, Decision Tree, Random Forest, Logistic Regression dan ANN, nilai akurasi dari algoritma pohon keputusan adalah 96,75% dan merupakan algoritma yang menurut algoritma Random Forest dengan akurasi 97,29% menawarkan akurasi tertinggi.(Safdari et al., 2022), penelitian kedua menggunakan Particle Swarm Optimization (PSO) menggunakan algoritma C.45 untuk memilih atribut akurasi penyakit hepatitis, dilakukan oleh Lis Saumi Ramdhani, menyiratkan bahwa menggunakan PSO meningkatkan hasil akurasi (Studi & Informatika, n.d.). Penelitian sebelumnya telah menemukan akurasi yang sangat baik, tetapi masih ada ruang untuk perbaikan. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi penyakit hepatitis C menggunakan pohon keputusan C4.5 dengan memilih fungsi PSO dan menganalisis hasilnya.

2. KERANGKA TEORI

2.1 Penelitian Terkait

Penelitian sebelumnya yang menjadi latar belakang penelitian ini dijabarkan pada Tabel 1 dibawah ini :

Tabel 1. Penelitian Terkait

No	Judul Dan Peneliti	Dataset	Metode	Hasil
1	Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Hepatitis Wisti Dwi Septiani	Machine Learning RepositoryUCI (Universitas California Invene) dengan	Klasifikasi data mining algoritma C.45	Akurasi 77, 29%

ISSN: 2598-0256, E-ISSN: 2598-0238

		alamat web: http://archive.ics.uci.edu/ml/		
2	Penerapan Particle Swarm Optimization (Pso) Untuk Seleksi Atribut Dalam Meningkatkan Akurasi Prediksi Diagnosis Penyakit Hepatitis Dengan Metode Algoritma C4.5 Lis Saumi Ramdhan	Database yang berasal dari http://archive.ics.uci.edu/ml/datasets/Hepatitis sebagai subset dari dataset publik yang digunakan dalam proyek statlog eropa	Data yang didapat dari UCI Machine Learning sebanyak 155 record, dari data tersebut terdapat data missing value sebanyak 5 record	Akurasi algoritma C4.5 senilai 79,33%, sedangkan untuk nilai akurasi Optimasi algoritma C4.5 menggunakan PSO sebesar 85,00%
3	Klasifikasi Hepatitis C Virus Menggunakan Algoritma C4.5 Classification Of Hepatitis C Virus Using Algorithm C4.5 Susanto dan Nuri	Data yang digunakan berasal dari UCI dataset	Menggunkan algoritma C4.5 dilakukan dengan menerapkan metode lain yaitu Metode Adaboost, akan meningkat,	Nilai akurasi yang dihasilkan dari Algritma C4.5 dengan Adaboost sebesar 95,60%

2.2. Hepatitis

Hepatitis merupakan suatu penyakit peradangan hati yang umumnya disebabkan oleh virus. Selain itu, hepatitis juga bisa disebabkan oleh alkohol dan penyakit autoimun. Hepatitis virus dapat timbul dari aktivitas yang terkontaminasi virus (misalnya penggunaan jarum suntik, obat suntik, jarum transfusi, jarum tato dan tindik, berhubungan seks dengan penderita hepatitis, atau berinteraksi dengan petugas terkait hepatitis). 5 jenis virus hepatitis yaitu A, B, C, D, kemudian E. Ciri-ciri dari masing-masing jenis ini berbeda-beda, sehingga gejala dan pengobatannya juga berbeda-beda(Latu Handarko, 2015).

Virus hepatitis telah menyebar ke seluruh dunia dan merupakan masalah kesehatan masyarakat global yang utama Tidak semua kasus hepatitis berkembang, tetapi gejala umum hepatitis termasuk demam, mual hingga muntah, lesu (mendengarkan), mudah memar, dan penyakit kuning (jaundice). Jika tidak diobati, hepatitis dapat berkembang menjadi sirosis (kerusakan hati permanen) dan akhirnya gagal hati. Tes darah adalah cara terbaik untuk memeriksa hepatitis, tetapi biopsi hati, yang menghilangkan sepotong kecil jaringan hati untuk pengujian laboratorium, juga dapat dilakukan. Selain itu, dokter dapat mendiagnosis hepatitis dengan melakukan pemeriksaan fisik terhadap gejala hepatitis, seperti kulit dan mata menguning. Riwayat kesehatan juga diperlukan untuk mengetahui di mana pasien terpapar virus hepatitis. Hepatitis dapat dicegah dengan menghindari faktor risiko penularan hepatitis dan dengan menerima imunisasi dan vaksinasi(Tinggi et al., 2022).

2.3. Data Mining

Data mining merupakan proses mengekstrak sejumlah besar data yang sebelumnya tidak diketahui (Adiba, 2021). Data mining juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dari database. Hal ini sering disebut sebagai penemuan pengetahuan dalam keputusan database (KDD) dan bertanggung jawab untuk penyebaran hasil. Pertumbuhan berkelanjutan dalam penambangan data dan penemuan pengetahuan didorong oleh beberapa penemuan (Daniel, 2005):

a. Pertumbuhan eksplosif dalam pengumpulan data sebagaimana dibuktikan oleh pemindaian supermarket
b. Menyimpan data di gudang data sehingga seluruh perusahaan memiliki akses ke stok data yang up-to-date dan otoritatif
c. Ketersediaan akses yang ditingkatkan ke data dari navigasi web dan intranet
d. Tekanan persaingan untuk meningkatkan pangsa pasar dalam perekonomian global
e. Pengembangan paket perangkat lunak penambangan data komersial siap pakai
f. Peningkatan daya komputasi dan kapasitas penyimpanan yang signifikan

ISSN: 2598-0256, E-ISSN: 2598-0238

2.4. Klasifikasi

Klasifikasi merupakan tatanan yang sangat penting dalam menambang data komunitas. Klasifikasi adalah teknik penambangan data prediktif yang menggunakan hasil yang diketahui dari kumpulan data yang berbeda untuk membuat prediksi tentang data nilai. Masalah dengan akurasi banyak algoritma klasifikasi adalah bahwa informasi diketahui hilang saat memproses data yang tidak seimbang, misalnya ketika distribusi sampel antar kelas sangat miring (Misdram et al., 2021). Dalam klasifikasi, Anda memiliki variabel target kategoris, seperti strata pendapatan, yang dapat, misalnya, membagi Anda menjadi tiga kelas atau kategori: pendapatan tinggi, pendapatan menengah, dan pendapatan rendah. Model penambangan data memeriksa kumpulan data besar. Setiap kumpulan data berisi informasi tentang variabel target dan satu set input atau variabel prediktor. Contoh tugas klasifikasi dalam bisnis dan penelitian meliputi (Daniel, n.d.):

a. Menentukan apakah transaksi kartu kredit tertentu adalah penipuan
b. Penempatan mahasiswa baru pada jalur khusus yang berkaitan dengan kebutuhan khusus
c. Mengevaluasi apakah aplikasi hipotek menimbulkan risiko kredit
d. Diagnosis adanya penyakit tertentu
e. Menentukan apakah wasiat itu ditulis oleh almarhum sendiri atau dipalsukan oleh orang lain
f. Menentukan Apakah Perilaku Keuangan atau Pribadi Tertentu Mengindikasikan Potensi Ancaman Teroris

Klasifikasi manual adalah klasifikasi yang dilakukan oleh manusia tanpa bantuan algoritma komputer cerdas. Ada beberapa algoritma untuk klasifikasi yang dilakukan menggunakan teknologi ini, antara lain naive Bayes, support vector machine, pohon keputusan, fuzzy, dan jaringan syaraf tiruan (Wibawa et al., 2018).

2.5. Decision Tree C.45

Pohon keputusan adalah salah satu jenis algoritma penambangan data yang paling populer untuk klasifikasi dan prediksi. Dtree mengatur catatan dalam struktur pohon yang terdiri dari simpul akar, cabang, dan simpul daun. Node akar berada di bagian atas struktur pohon. Node mewakili atribut, cabang mewakili hasil, lalu daun mewakili keputusan. (Khomsah, n.d.).

Ada beberapa langkah untuk membangun pohon keputusan menggunakan algoritma C4.5 (Tinggi et al., 2022) yaitu:

a. Memerlukan pelatihan data, dapat diambil dari data historis yang telah terjadi sebelumnya dan dikelompokkan ke dalam kelas-kelas tertentu.
b. b. Tentukan akar pohon dengan menghitung nilai gain tertinggi dari setiap atribut atau nilai indeks entropi terendah. Sebelumnya, nilai indeks entropi dihitung menggunakan rumus:

$$Entropy\ (i) = \sum_{j=1}^{m} f(i,j).\ 2f[(i,j)] \qquad (1)$$

c. Nilai gain dengan rumus:

$$gain = -\sum_{i=1} \frac{ni}{n} .IE(i) \qquad (2)$$

d. Untuk menghitung gain ratio perlu diketahui suatu term baru yang disebut Split Information dengan rumus:

$$SplitInformation = -\sum_{t=1} \frac{S1}{S} \log 2 \frac{S1}{S} \qquad (3)$$

e. Selanjutnya menghitung gain ratio

$$Gainratio(S,A) = \frac{Gain\ (S.A)}{SplitInformation\ (S,A)} \qquad (4)$$

f. Ulangi langkah 2 sampai semua record telah terpecah. Proses pemisahan pohon keputusan berakhir ketika:
   1) Semua tupel dalam catatan simpul m adalah kelas yang sama.
   2) Atribut dalam dataset tidak dibagi lagi.
   3) Cabang kosong tidak memiliki catatan

Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022
Institut Informatika dan Bisnis Darmajaya, 27 Agustus 2022
ISSN: 2598-0256, E-ISSN: 2598-0238

2.6. Particle Swarm Optimization

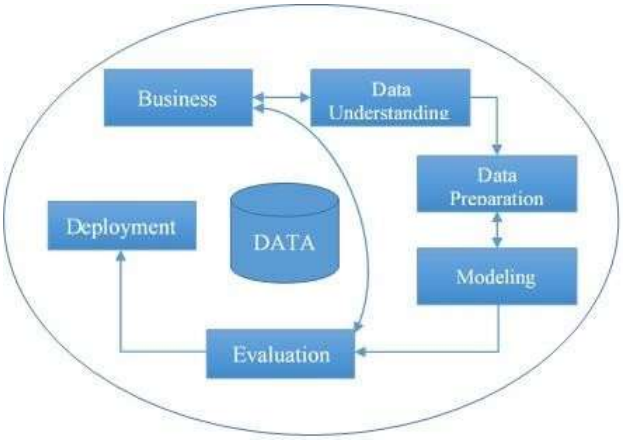
Particle Swarm Optimization (PSO) adalah teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Particle Swarm Optimization (PSO) memiliki beberapa teknik untuk optimasi, seperti meningkatkan bobot atribut untuk atribut atau variabel yang digunakan, memilih atribut (attribute selection), dan seleksi fitur(Mustopa, 2021). Particle swarm Optimization adalah algoritma yang terinspirasi oleh perilaku sosial hewan seperti burung, lebah, dan ikan. Hewan dalam algoritma PSO dianggap partikel. Partikel ini tunduk pada kecerdasan individu hewan itu sendiri dan kecerdasan partikel lain di dalam kelompok. Jika suatu partikel menemukan jalur terpendek yang benar ke sumber makanan, partikel lain akan mengikuti partikel yang sebelumnya menemukan jalur terpendek yang benar (Hakim et al., 2017)

2.7. Confusion Matrix

Matriks konfigurasi adalah tabel yang terdiri dari jumlah baris data uji yang diprediksi benar dan salah dengan menggunakan model klasifikasi yang digunakan. Tabel matriks kebingungan diperlukan untuk memilih model klasifikasi yang berkinerja terbaik (Romadhon & Kurniawan, 2021). Confusion matrix adalah matriks 2x2 yang merepresentasikan hasil klasifikasi biner dalam suatu kumpulan data. Ada beberapa rumus umum yang dapat digunakan untuk menghitung kinerja klasifikasi. Hasil prosentase untuk nilai Accuracy, Precision dan Recall dapat ditampilkan (Andika et al., 2019)

2.8. Metode CRISP–DM

CRISP menyediakan proses standar nonproprietary dan tersedia secara bebas untuk memasukkan data mining ke dalam strategi pemecahan masalah umum dari bisnis atau unit penelitian. Menurut CRISP-DM, sebuah proyek data mining tertentu memiliki siklus hidup yang terdiri dari enam fase, seperti yang ditunjukkan pada Gambar 1. Perhatikan bahwa urutan fase adaptif. Dengan kata lain, fase berikutnya dalam urutan sering tergantung pada hasil yang terkait dengan fase sebelumnya. Ketergantungan yang paling penting antara fase ditunjukkan oleh panah. Misalkan Anda berada dalam tahap pemodelan. Tergantung pada perilaku dan properti model, mungkin perlu kembali ke fase persiapan data untuk penyempurnaan lebih lanjut sebelum melanjutkan ke fase evaluasi model. Sifat iteratif CRISP diwakili oleh lingkaran luar pada gambar.



Gambar. 1 Tahapan CRISP-DM

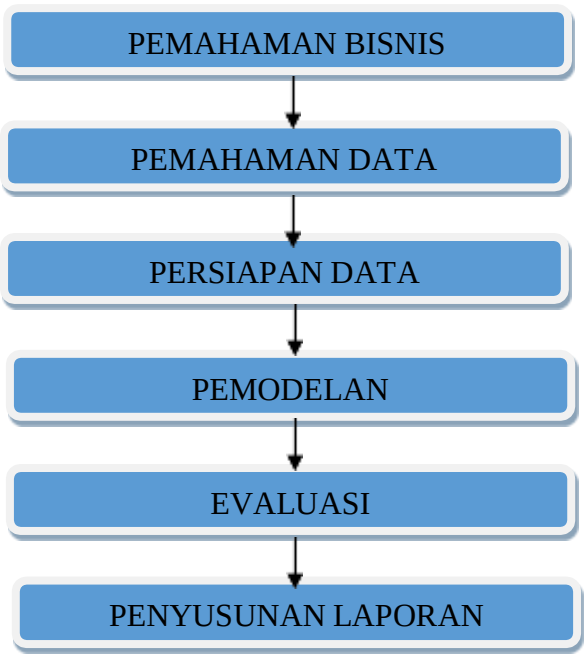
Solusi untuk masalah bisnis atau penelitian tertentu sering kali menimbulkan pertanyaan menarik lainnya yang dapat diatasi dengan menggunakan proses umum yang sama seperti sebelumnya. Temuan dari proyek-proyek masa lalu harus selalu digunakan sebagai masukan untuk proyek-proyek baru. Di bawah ini adalah gambaran dari setiap fase. Jika masalah muncul selama tahap evaluasi, analis dapat dikirim kembali ke salah satu tahap sebelumnya untuk perbaikan, tetapi untuk kesederhanaan, hanya loop paling umum yang ditampilkan dan tahap pemodelan dikembalikan (Daniel, 2005).

ISSN: 2598-0256, E-ISSN: 2598-0238

3. METODOLOGI

3.1 Tahapan Data Mining

Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur dalam tahapan penelitian yang akan dilakukan berikut:



Gambar. 2 Tahapan Data Mining

Tahapan pada Gambar 2 merupakan proses data mining dalam penelitian ini.

1. Fase pemahaman bisnis, juga dikenal sebagai fase pemahaman penelitian, berisi tujuan dan persyaratan proyek yang jelas yang berhubungan dengan seluruh perusahaan atau entitas penelitian, mengubah tujuan dan kendala menjadi perumusan definisi masalah penambangan data, Mempersiapkan Strategi untuk mencapai tujuan tersebut tercapai.
2. Fase pemahaman data dengan mengumpulkan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data untuk mendapatkan wawasan awal, menilai kualitas data, dan memilih subset yang menarik jika sesuai.
3. Fase persiapan data menyiapkan data mentah awal untuk kumpulan data akhir yang digunakan di semua fase berikutnya. Pada tahap ini, kami memilih kasus dan variabel yang kami butuhkan dan menganalisis yang cocok untuk analisis kami. Jika perlu, melakukan transformasi pada variabel tertentu akan membersihkan data mentah dan membuatnya siap untuk alat pemodelan.
4. Fase Pemodelan Pada fase ini, teknik pemodelan yang tepat dipilih dan diterapkan. Kalibrasi pengaturan model untuk mengoptimalkan hasil. Seringkali ada beberapa teknik berbeda yang tersedia untuk masalah data mining yang sama. Jika perlu, kembali ke fase persiapan data untuk menyesuaikan bentuk data sesuai dengan kebutuhan spesifik teknik penambangan data Anda.
5. Selama Tahap Evaluasi, model atau model yang diajukan selama Tahap Modeling dievaluasi kualitas dan efektivitasnya sebelum dikerahkan untuk digunakan di lapangan, menentukan tercapai atau tidaknya tujuan yang ditetapkan. aspek kunci dari bisnis atau masalah penelitian belum ditangani secara memadai dan menggunakan hasil penambangan data
6. Tahapan penyusunan laporan disertasi adalah mendokumentasikan apa yang relevan dengan pekerjaan penelitian yang dilakukan. Mencatat hasil penelitian dan menerjemahkannya ke dalam argumentasi yang disajikan dalam bentuk laporan dan dapat digunakan sebagai literatur.

Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022
Institut Informatika dan Bisnis Darmajaya, 27 Agustus 2022
ISSN: 2598-0256, E-ISSN: 2598-0238

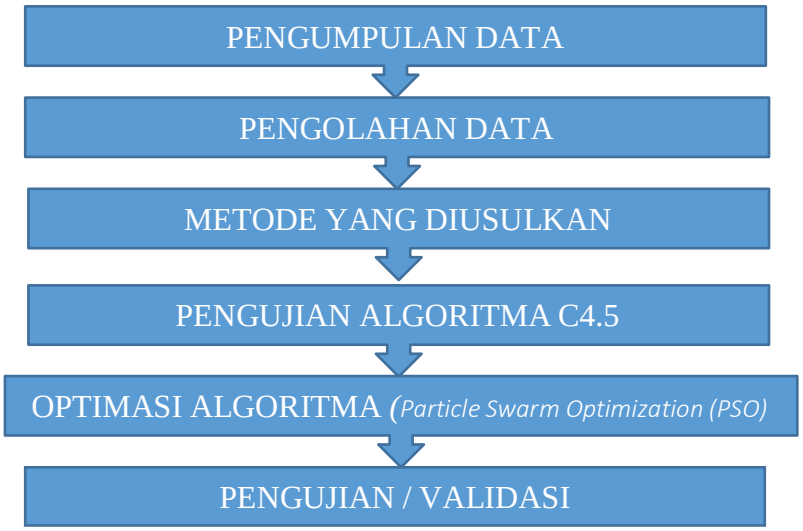
3.2 Kerangka Pemikiran

<table>
<tr><th colspan="3">MASALAH</th></tr>
<tr><td colspan="3">Hasil yang didapatkan pada percobaan sebelumnya menggunakan algoritma C4.5 lebih rendah dibandingkan Algoritma Particle Swarm Optimization (PSO) maka perlu adanya peningkatan pada algoritma C4.5</td></tr>
<tr><th colspan="3">STUDI LITERLATUR</th></tr>
<tr><td colspan="3">Mencari Referensi-referensi teori yang sesuai dengan permasalahan yang akan dilakukannya penelitian.</td></tr>
<tr><th colspan="3">PERCOBAAN</th></tr>
<tr><td>Data</td><td>Metode</td><td>Tools</td></tr>
<tr><td>Dataset Hepatitis C</td><td>Algoritma C4.5</td><td>Rapidminer</td></tr>
<tr><td></td><td>Algoritma Particle Swarm Optimization (PSO)</td><td></td></tr>
<tr><th colspan="3">PENGUKURAN</th></tr>
<tr><td colspan="3">Acccuracy,Precision, Recall, Performance Vektor</td></tr>
<tr><th colspan="3">HASIL</th></tr>
<tr><td colspan="3">Teknk optimasi menggunakan algoritma PSO berhasil meningkatkan akurasi dari algoritma C4.5 dalam penentuan hepatitis C. Hasil prediksi hepatitis C dari algoritma C4.5 tanpa optimasi adalah sebesar 99,35 % dan setelah dioptimasi dengan algoritma PSO adalah sebesar 99,67 % maka ada peningkatan akurasi sebesar 0,32 %</td></tr>
</table>

Gambar. 3 Kerangka Pemikiran

Dalam penelitian ini perlu adanya kerangka pemikiran yang digunakan untuk sebagai landasan serta pedoman agar penelitian ini berjalan sesuai dengan alur yang direncanakan. Permasalahan pada penelitian ini adalah belum adanya metode yang dapat digunakan untuk mengoptimasi data hepatitis C. Metode yang digunakan pada penelitian ini adalah algoritma C4.5 Decision tree dan juga algoritma *Particle Swarm Optimization (PSO*) untuk dilakukan pengujian. Pengujian dari metode yang telah diterapkan menggunakan cara *Confusion Matrix* dan *Accuracy Performance Vector (Performance).* Untuk *tool* yang digunakan untuk melakukan pengujian metode adalah aplikasi RapidMiner.

3.3 Tahapan Penelitian



Gambar. 4 Tahapan Penelitian

ISSN: 2598-0256, E-ISSN: 2598-0238

Adapun tahapan penelitian yang akan dilakukan sebagai berikut :

1) Pengumpulan Data

   Pada bagian pengumpulan data dijelaskan dari mana data dalam penelitian ini didapatkan, meliputi data kuantitatif dan data kualitatif. Data kuantitatif yang diperoleh dari sumber perusahaan Percetakan untuk keperluan penelitian, sedangkan data kualitatif berisi tentang data yang dihasilkan dari penelitian.

2) Pengolahan Awal Data

   Pengelompokan awal data menjelaskan tentang tahapan awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, dalam pengelompokan dan penentuan atribut data.

3) Metode yang Diusulkan

   Metode yang Diusulkan menjelaskan tentang metode yang diusulkan untuk mengoptimasi klasifikasi penjualan barang yang laris dan kurang laris. Penjelasan ini meliputi pengaturan dan pemilihan dari atribut-atribut yang digunakan sebagai parameter dan arsitektur melalui uji coba.

4) Eksperimen dan Pengujian Metode

   Pada bagian Eksperimen dan Pengujian Metode menjelaskan tentang langkah-langkah eksperimen meliputi cara pemilihan arsitektur yang tepat dari model atau metode yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat.

5) Evaluasi dan Validasi Hasil

   Pada bagian Evaluasi dan Validasi Hasil dijelaskan tentang evaluasi dan validasi hasil penerapan metode pada penelitian yang dilakukan.

3.4 Pengolahan Data Awal

Data tersebut berisi 615 observasi dan 14 atribut laboratorium dan nilai demografis pendonor darah dan pasien Hepatitis C data bersumber dari www.kaggle.com ((https://www.kaggle.com/datasets/amritpal333/hepatitis-c-virus-blood=biomarkers). Sekuruh atribut pada data kecuali Kategori dan Jenis Kelamin adalah numerik. Atribut 1 sd 4 merujuk pada data pasien dan 5 sampai dengan 14 adalah hasil cek darah laboratorium. Target atau label data yang digunakan adalah Category atribut yang digunakan untuk perhitungan 13 atribut data.

Tabel 2. Atribut Dataset

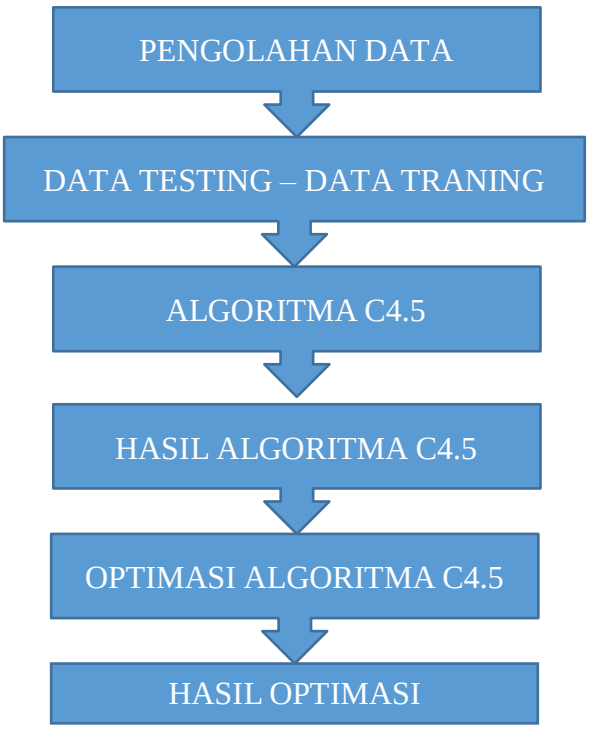
No.	Atribut	Skala Pengukuran
1	Att1 (Pasien ID/No.)	[1-615]
2	Category (Diagnosa)	0=Blood Donor (0=donor darah) 3=Cirrhosis (penyakit hati stadium akhir) 2=Fibrosis (jaringan parut fibrotik) 1=Hepatitis (1=peradangan hati) 0s=Suspect Blood Donor (0s=dugaan Donor Darah)
3	*Age* (Usia)	[19-77]
4	*Sex* (Jenis Kelamin)	[m = Male f = Female]
5	ALB (jumlah albumin)	[14,9-82,2]
6	ALP (jumlah alkaline phosphatase)	[19,1-190,7]

Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022
Institut Informatika dan Bisnis Darmajaya, 27 Agustus 2022
ISSN: 2598-0256, E-ISSN: 2598-0238

7	ALT (jumlah alanin transaminase)	[0,9-97,8]
8	AST (jumlah aspartat aminotransferase)	[10,6-114,7]
9	BIL (jumlah bilirubin)	[0,8-91]
10	CHE (jumlah kolinesterase)	[1,42-16,41]
11	CHOL (jumlah kolesterol)	[1,43-9,67]
12	CREA (jumlah kreatin)	[9-1079,1]
13	GGT (jumlah gamma-glutamil transferase)	[4,5-158,2]
14	PROT (jumlah protein)	[44,8-90]

3.5 Desain Penelitian

Dalam penelitian ini melalui tahapan dilakukannya uji coba menggunakan metode algoritma C4.5 dan Optimasi *Particle Swarm Optimization (PSO*). Data dianalisa dengan menggunakan algoritma sesuai dengan metode yang telah di rencanakan, setelah itu membandingkan metode dengan optimasi *Particle Swarm Optimization (PSO)* untuk melihat perbandingan tertinggi tingkat akurasi. Pada tahapan uji coba ini akan dilakukan beberapa langkah atau tahapan pengujian data yaitu seperti berikut:



Gambar. 5 Pengujian Data

Dari gambar metode yang diusulkan akan mendapat dan mengetahui hasil yang didapatkan oleh algoritma C4.5, kemudian pada hasil tersebut akan ditambahkan algoritma *Particle Swarm Optimization (PSO*) sebagai optimasi yang digunakan untuk meningkatkan hasil dari algoritma C4.5. Pada tahap akhir hasil dari keduanya akan dibandingkan sehingga akan diketahui hasil dari uji coba seberapa efektifnya tingkat optimasi algoritma *Particle Swarm Optimization (PSO*) pada dataset Hepatitis C.

ISSN: 2598-0256, E-ISSN: 2598-0238

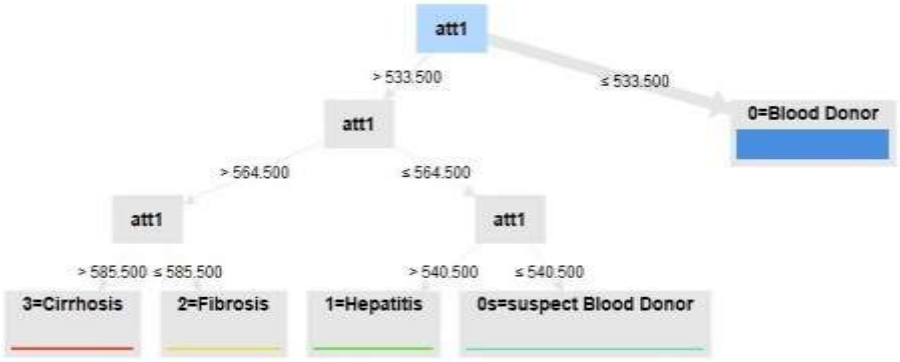
4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menerapkan dan membandingkan optimasi swarm partikel untuk meningkatkan akurasi prediksi hepatitis C menggunakan metode C4.5. Hasil dari penelitian ini sendiri berupa hasil pengolahan data kuantitatif dan kuantitatif dengan perhitungan berdasarkan model yang diajukan. Studi ini dilakukan pada dataset yang tersedia untuk umum. Eksperimen dan pengujian pada penelitian ini dilakukan dengan metode C4.5 tanpa PSO kemudian dengan metode PSO.

4.1 Eksperimen dan pengujian model C4.5

Pengelolaan data menggunakan model C4.5 dilakukan pada dataset yang terdiri dari 14 atribut yang merupakan atribut dari diagnosis penyakit Hepatitis C dan class yang merupakan hasil akhir prediksi. Model dari algoritma C4.5 yaitu berupa pohon keputusan, agar lebih mudah dalam membuat pohon keputusan, langkah pertama adalah menghitung jumlah *class* yang berpotensi '0=Donor Darah', '0s=Dugaan Donor Darah', '1=Hepatitis', '2=Fibrosis', '3=Sirosis' dari masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*. Dalam menyeleksi atribut dibutuhkan suatu algoritma *particle swarm optimization* pada C4.5, untuk proses pembelajarannya dieksperimenkan pada Rapidminer. Sedangkan proses pengujiannya menggunakan *Accuracy* dan *Performace* untuk melihat tingkat keakuratan prediksi diagnosis penyakit hepatitis C. Atribut yang digunakan yaitu (Pasien ID/No), Usia, Jenis Kelamin ('f'=perempuan, 'm'=laki-laki), ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, PROT), Kategori (Fitur target. nilai: '0=Donor Darah', '0s=Dugaan Donor Darah', '1=Hepatitis', '2=Fibrosis', '3=Sirosis'). Dari 14 atribut yang terdapat pada dataset *kaggle*, kemudian selanjutnya diseleksi menjadi hanya 5 atribut yang digunakan dalam menentukan prediksi penyakit hepatitis C, atribut-atribut tersebut yaitu : (Pasien ID/No), Usia, AST, BIL, CHE, CREA), Kategori (Fitur target. nilai: '0=Donor Darah', '0s=Dugaan Donor Darah', '1=Hepatitis', '2=Fibrosis', '3=Sirosis'). Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *particle swarm optimization* mampu menyeleksi atribut pada C4.5, sehingga menghasilkan tingkat akurasi diagnosis penyakit hepatitis C yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5.

Dari proses yang telah dijalankan dalam penjelasan berikut maka didapatkan hasil sebagai berikut :



Gambar. 6 Pohon Keputusan Decision Tree C4.5

Berikut adalah hasil accuracy, recall, precision.

accuracy: 99.35% +/- 1.14% (micro average: 99.35%)

	true 0=Blood Donor	true 0s=suspect Blood...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood Donor	533	2	0	0	0	99.63%
pred. 0s=suspect Bloo...	0	5	1	0	0	83.33%
pred. 1=Hepatitis	0	0	23	0	0	100.00%
pred. 2=Fibrosis	0	0	0	21	1	95.45%
pred. 3=Cirrhosis	0	0	0	0	29	100.00%
class recall	100.00%	71.43%	95.83%	100.00%	96.67%	

Gambar. 7 Accuracy Decision Tree C.45

Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2022
Institut Informatika dan Bisnis Darmajaya, 27 Agustus 2022
ISSN: 2598-0256, E-ISSN: 2598-0238

Dari Gambar diatas maka dapat dilihat bahwa nilai *accuracy* yang didapat daripengujian algoritma *Decision Tree* C4.5 yaitu sebesar 99,35 %. Dan nilai hasil dari *recall, precision* dapat dilihat pada gambar di atas.

accuracy: 99.67% +/- 0.69% (micro average: 99.67%)

	true 0=Blood Donor	true 0s=suspect Blood...	true 1=Hepatitis	true 2=Fibrosis	true 3=Cirrhosis	class precision
pred. 0=Blood Donor	533	1	0	0	0	99.81%
pred. 0s=suspect Bloo...	0	6	0	0	0	100.00%
pred. 1=Hepatitis	0	0	24	1	0	96.00%
pred. 2=Fibrosis	0	0	0	20	0	100.00%
pred. 3=Cirrhosis	0	0	0	0	30	100.00%
class recall	100.00%	85.71%	100.00%	95.24%	100.00%	

Gambar. 8 Accuracy Decision Tree C.45 + PSO

Maka dapat dilihat bahwa nilai *accuracy* yang didapat dari pengujian algoritma *Decision Tree* C4.5 + Algoritma PSO yaitu sebesar 99,67 %. Dan nilai hasil dari *recall, precision* dapat dilihat pada gambar di atas.

4.2 Pembanding Pengujian Model Algoritma C4.5 dengan Algoritma C4.5 – PSO

Tabel.3 Pembanding Pengujian Sebelum dan Seseudah Optimasi

Algoritma	Accuracy	Recall	Precesion	Prediksi	
C.45	99,35 %	100.00%	99.63%	553	0=Blood Donor
		71.43%	83.33%	5	0s=suspect Blood Donor
		95.83%	100.00%	23	1=Hepatitis
		100.00%	95.24%	20	2=Fibrosis
		96.67%	100.00%	29	3=Cirrhosis
C.45 -PSO	99,67 %	100 %	99.81%	553	0=Blood Donor
		85.71%	100.00%	6	0s=suspect Blood Donor
		100.00%	95%	24	1=Hepatitis
		95.24%	100.00%	21	2=Fibrosis
		100.00%	100.00%	30	3=Cirrhosis

5. KESIMPULAN

Dari hasil penelitian selama ini dapat disimpulkan bahwa pengolahan data pada pengujian ini adalah algoritma pohon keputusan C4.5. Dan hasil yang didapat saat pengujian data hepatitis dengan kategori *0=donor darah, 0s=dugaan donor darah, 1=hepatitis, 2=fibrosis, 3=sirosis) menggunakan metode algoritma decision tree C4.5 hanya di Classification untuk mendapatkan nilai akurasinya. Belut yang optimal adalah 99,35%. Pengujian selanjutnya menggunakan algoritma pohon keputusan C4.5 dengan optimasi algoritma PSO memberikan nilai akurasi optimal sebesar 99,67%, sehingga optimasi dengan algoritma PSO mengurangi desil algoritma akurasi keputusan aklegaritha silkanehakhini silnah 4.5 sebesar 0,32 % dapat meningkat.

UCAPAN TERIMA KASIH

Terima kasih keluarga yang tak pernah henti memberikan support, teman-temen seperjuangan MTI angkatan 22 IIB Darmajaya.

ISSN: 2598-0256, E-ISSN: 2598-0238

DAFTAR PUSTAKA

Adiba, F. (2021). Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes. *Indonesian Journal of Fundamental Scienes*, *7*(1), 18–28.

Alhawaris. (2019). Hepatitis C: Epidemiologi, Etiologi, dan Patogenitas. *Jurnal Sains Dan Kesehatan*, *2*(2), 139–150.https://doi.org/10.25026/jsk.v2i2.132

Andika, L. A., Amalia, P., & Azizah, N. (2019). *Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier*. *2*(1), 34–41.

Daniel, T. (2005). *Discovering Knowledge in Data*.

Hakim, S. H. F., Cholissodin, I., & Widodo, A. W. (2017). Seleksi Fitur Dengan Particle Swarm Optimization Untuk Pengenalan Pola Wajah Menggunakan Naive Bayes (Studi Kasus Pada Mahasiswa Universitas Brawijaya Fakultas Ilmu Komputer Gedung A). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(10), 1045–1057.

Khomsah, S. (n.d.). *Prediksi Harapan Hidup Penderita Hepatitis Kronik Menggunakan Metode-Metode Klasifikasi*.

Lanini, S., Easterbrook, P. J., Zumla, A., & Ippolito, G. (2016). Hepatitis C: global epidemiology and strategies for control. In *Clinical Microbiology and Infection* (Vol. 22, Issue 10, pp. 833–838). Elsevier B.V. https://doi.org/10.1016/j.cmi.2016.07.035

Latu Handarko, J. (2015). Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis. *UJM*, *4*(2).http://journal.unnes.ac.id/sju/index.php/ujm

Milovic, B., & Milovic, M. (2012). Prediction and Decision Making in Health Care using Data Mining. *International Journal of Public Health Science (IJPHS)*, *1*(2), 69–78.

Mustopa, A. (2021). *Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization*. *2*.

Safdari, R., Deghatipour, A., Gholamzadeh, M., & Maghooli, K. (2022). Applying data mining techniques to classify patients with suspected hepatitis C virus infection. *Intelligent Medicine*. https://doi.org/10.1016/j.imed.2021.12.003

Studi, P., & Informatika, M. (n.d.). *Lis Saumi Ramdhani*.

Tinggi, S., Pati, T., Korespondensi, P., & Virus, H. C. (2022). *KLASIFIKASI HEPATITIS C VIRUS MENGGUNAKAN ALGORITMA C4 . 5 CLASSIFICATION OF HEPATITIS C VIRUS USING ALGORITHM C4 . 5*. *13*(2), 43–48. https://doi.org/10.34001/jdpt.v12i2

WHO. (2022). *Hepatitis C*. https://www.who.int/news-room/fact-sheets/detail/hepatitis-c. 5/08/2022