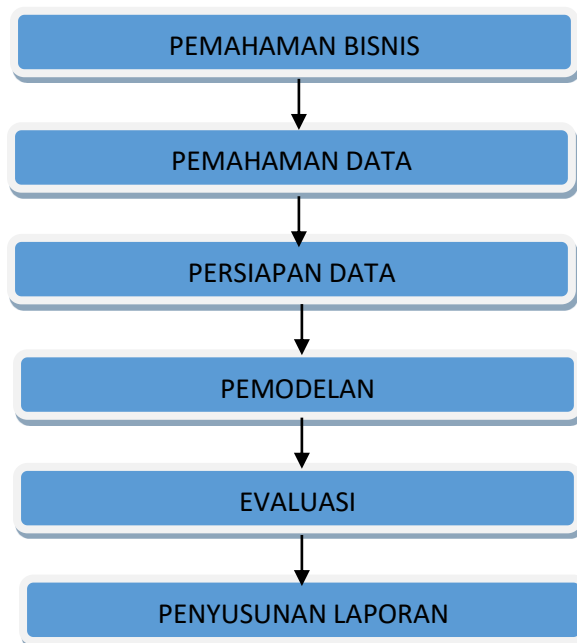


## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

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. 3.1 Tahapan Data Mining*

Tahapan pada gambar 2 Adalah proses data mining pada penelitian ini :

1. Tahap Pemahaman bisnis juga dapat disebut fase pemahaman penelitian yang meliputi tujuan dan persyaratan proyek dengan jelas dalam hal bisnis atau unit penelitian secara keseluruhan, menterjemahkan tujuan dan batasan ke dalam perumusan definisi masalah data mining, menyiapkan strategi awal untuk mencapai tujuan tersebut.
2. Tahap pemahaman data dengan melakukan pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal, mengevaluasi kualitas data, apabila diinginkan pilih subset menarik yang mungkin berisi pola yang dapat ditindaklanjuti.
3. Tahap persiapan data adalah mempersiapkan data mentah awal kumpulan data akhir yang akan digunakan untuk semua tahapan berikutnya. Pada tahapan ini

memilih kasus dan variabel yang diinginkan, menganalisis yang sesuai untuk analisis. Melakukan transformasi pada variabel tertentu, jika diperlukan, membersihkan data mentah sehingga siap untuk alat pemodelan.

Data tersebut berisi 615 observasi dan 14 atribut laboratorium dan nilai demografis pendonor darah dan pasien Hepatitis C data bersumber dari (<https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset?resource=download&select=HepatitisCdata.csv>). Seluruh atribut pada data kecuali Kategori dan Jenis Kelamin adalah nominal dan numerik. Atribut 1 sd 3 merujuk pada data pasien, 4 sampai dengan 13 adalah hasil cek darah laboratorium. Target atau label data yang digunakan adalah Category atribut yang digunakan untuk perhitungan 12 atribut data.

Tabel 3.1 Atribut-atribut Dataset

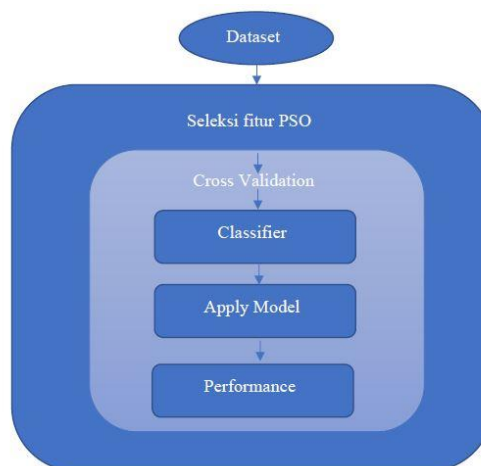
No.	Atribut	Tipe	Keterangan
1	Att1	integer	(Pasien ID/No.)
2	Age	integer	(Usia)
3	Sex	polynomial	(Jenis Kelamin)
4	ALB	polynomial	(jumlah albumin)
5	ALP	polynomial	(jumlah alkaline phosphatase)
6	ALT	polynomial	(jumlah alanin transaminase)
7	AST	polynomial	(jumlah aspartat aminotransferase)
8	BIL	polynomial	(jumlah bilirubin)
9	CHE	polynomial	(jumlah kolinesterase)
10	CHOL	polynomial	(jumlah kolesterol)
11	CREA	polynomial	(jumlah kreatin)
12	GGT	polynomial	(jumlah gamma- lutamil transferase)
13	PROT	polynomial	(jumlah protein)

Tabel 3.1 Atribut-atribut Dataset

14	Category (Diagnosa)	polynomial	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)

Tabel diatas menjelaskan masing-masing atribut dari data pada penelitian ini. Tujuan dari dataset ini adalah untuk menentukan diagnosa penyakit hepatitis C.

4. Tahap pemodelan, Tahapan ini dilakukan dengan memilih dan menerapkan teknik pemodelan yang sesuai. Kalibrasi pengaturan model untuk mengoptimalkan hasil, seringkali, beberapa teknik berbeda dapat digunakan untuk masalah data mining yang sama, apabila diperlukan mengulangi kembali ke tahapan persiapan data untuk membawa bentuk data sesuai dengan persyaratan khusus dari teknik data mining tertentu.



Gambar. 1.2 Tahapan Pemodelan

5. Tahap evaluasi adalah mengevaluasi satu atau lebih model yang disampaikan dalam tahap pemodelan untuk kualitas dan efektivitas sebelum menerapkannya untuk digunakan di lapangan.. Pada tahap ini yang akan dilakukan adalah menentukan model sebenarnya mencapai tujuan yang ditetapkan untuk itu pada tahapan pertama, menetapkan apakah beberapa aspek penting dari masalah bisnis atau penelitian belum diperhitungkan secara memadai, mengambil keputusan tentang

penggunaan hasil data mining menggunakan tools RapidMiner. Dengan tools tersebut akan dilakukan pengujian keakurasian dengan beberapa teknik yaitu confusion matrix

6. Tahap penyusunan laporan adalah mendokumentasikan yang berkaitan dengan penelitian yang dilakukan. Pencatatan hasil penelitian dan diterjemahkan dalam pembahasan untuk dapat disajikan dalam bentuk laporan dan dapat dijadikan literatur.

### 3.2 Kerangka Pemikiran

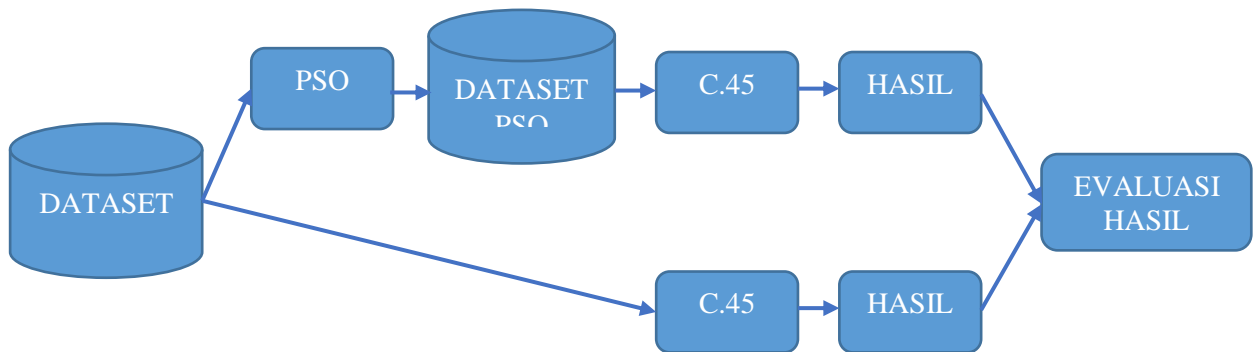
MASALAH		
Hasil yang didapatkan pada percobaan sebelumnya menggunakan algoritma C4.5 lebih rendah dibandingkan Algoritma <i>Particle Swarm Optimization (PSO)</i> maka perlu adanya peningkatan pada algoritma C4.5		
STUDI LITERATUR		
Mencari Referensi-referensi teori yang sesuai dengan permasalahan yang akan dilakukannya penelitian.		
PERCOBAAN		
Data	Metode	Tools
Dataset Hepatitis C	Algoritma C4.5	Rapidminer
	Algoritma C.45+ <i>Particle Swarm Optimization (PSO)</i>	
PENGUKURAN		
Accuracy, Precision, Recall, Performance Vektor		
HASIL		
Teknik optimasi menggunakan algoritma PSO berhasil meningkatkan akurasi dari algoritma C4.5 dalam penentuan hepatitis C.		

Gambar. 3.3 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 Desain Penelitian

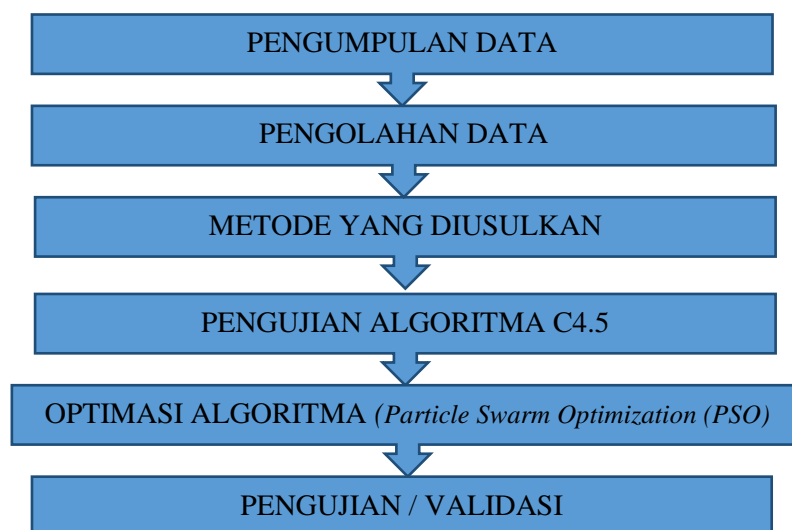
Dalam penelitian ini melalui tahapan dilakukannya uji coba menggunakan metode algoritma C4.5 dan Optimasi *Particle Swarm Optimization (PSO)*. Data dianalisa dengan menggunakan algoritma sesuai dengan metode yang telah di rencanakan, setelah itu membandingkan metode dengan optimasi *Particle Swarm Optimization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



Gambar. 3.4 Skema Riset

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.

### 3.4 Tahapan Penelitian



Gambar. 3.5 Tahapan Penelitian

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 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 penyakit hepatitis. 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.

### 3.5 Alat dan Bahan

Penelitian ini menggunakan perangkat keras Laptop Lenovo Thinkpad X260 Intel® Core™ i5 CPU 2.50 Ghz, RAM 8 GB sedangkan perangkat lunak yang digunakan Windows 10 Pro, Aplikasi RapidMiner Studio9.10.001 yang digunakan untuk pengolahan data.