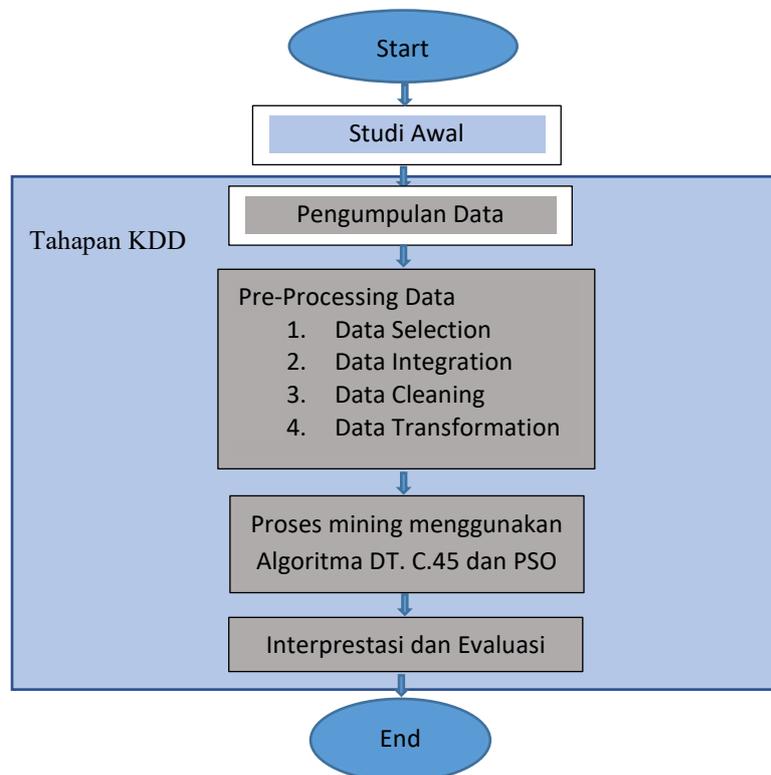


BAB III METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pada Bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, 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 Alur Dalam Tahapan Penelitian

Berikut ini merupakan penjelasan dari gambar 3.1 yang merupakan alur dari penelitian ini

1. Studi Awal.

Studi awal merupakan proses mencari dan mempelajari masalah yang akan dijadikan topik penelitian.

2. Pengumpulan Data

Data yang digunakan berasal dari dataset publik heart disease yang di peroleh dari <https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28original%29>. Dengan jumlah 699 record data, dengan 11 atribut di mana dari 10 atribut tersebut terdapat 1 atribut Class / informasi dan terdapat 16 record error dari 699 record menjadi 683 record, sehingga dataset yang diolah berjumlah 683 record.

Tabel 3.1 Dataset

Sample code number	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Class
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2
1033078	4	2	1	1	2	1	2	1	1	2
1035283	1	1	1	1	1	1	3	1	1	2
1036172	2	1	1	1	2	1	2	1	1	2
1041801	5	3	3	3	2	3	4	4	1	4
1043999	1	1	1	1	2	3	3	1	1	2
1044572	8	7	5	10	7	9	5	5	4	4
1047630	7	4	6	4	6	1	4	3	1	4
1048672	4	1	1	1	2	1	2	1	1	2
1049815	4	1	1	1	2	1	3	1	1	2
1050670	10	7	7	6	4	10	4	1	2	4
1050718	6	1	1	1	2	1	3	1	1	2
1054590	7	3	2	10	5	10	5	4	4	4
1054593	10	5	5	3	6	7	7	10	1	4
1056784	3	1	1	1	2	1	2	1	1	2

Dari dataset diatas terdapat label yang diberi nama class dimana dari data class di atas terdapat 2 tipe yaitu Jinak dan ganas payudara, yaitu:

1. Sebanyak 239 di kategorikan tumor jinak payudara (*Benign Breast Tumors*) dengan kode angka class 2
2. Sebanyak 444 di kategorikan tumor ganas payudara (*Malignant Breast Tumor/Breast Cancer*) dengan kode class angka 4

Data yang digunakan dalam penelitian ini adalah data sampel dari public dataset UCI, karena data dari pasien kanker payudara merupakan data yang private dan sulit untuk didapatkan.

Tabel 3.2 Keterangan Atribut Dataset

N0	Attribute	Domain	Tipe	Keterangan
1	Sample code number	id number	Integer	Contoh nomor kode
2	Clump Thickness	1-10	Integer	Ketebalan rumpun
3	Uniformity of Cell Size	1-10	Integer	Keseragaman Ukuran Sel
4	Uniformity of Cell Shape	1-10	Integer	Keseragaman Bentuk Sel
5	Marginal Adhesion	1-10	Integer	Adhesi Marjinal
6	Single Epithelial Cell Size	1-10	Integer	Ukuran Sel Epitel Tunggal
7	Bare Nuclei	1-10	Integer	Ukuran Asli Nuclei
8	Bland Chromatin	1-10	Integer	Kromatin
9	Normal Nucleoli	1-10	Integer	Keadaan Nucleoli Normal
10	Mitoses	1-10	Integer	Mitosis
11	Class:	2(benign) 4 (malignant)	Binominal	Kelas : (2 untuk Jinak, 4 untuk Ganas)

Parameter-parameter kanker payudara terdiri dari sebagai berikut menurut Frank,

A. dan Asuncion dalam (Hermawanti, 2015):

1. *Clump thickness* sel *benign* cenderung dikelompokkan dalam *monolayers*, sementara sel-sel kanker sering dikelompokkan dalam *multilayers*.
2. *Uniformity of cell size* sel-sel kanker mempunyai ukuran bervariasi.
3. *Uniformity of cell shap* sel-sel kanker mempunyai bentuk bervariasi.
4. *Marginal adhesion* sel-sel normal cenderung tetap bersama-sama.

5. *Single epithelial cell size* sel-sel epitel yang signifikan diperbesar menjadi sel *malignant*.
6. *Bare nuclei* adalah istilah yang digunakan untuk inti (*nuclei*) yang tidak dikelilingi oleh *cytoplasm* (seluruh sel). Biasanya terlihat di benign.
7. *Bland Chromatin* inti “tekstur” seragam yang dilihat dalam sel benign. Dalam sel-sel kanker chromatin cenderung lebih kasar.
8. *Normal nucleoli* adalah struktur kecil yang terlihat dalam inti atom, pada sel-sel normal nucleolus biasanya sangat kecil jika terlihat sama sekali dalam sel-sel kanker *nucleoli* menjadi lebih menonjol.
9. *Mitoses* : pembelahan satu sel menjadi dua sel.
10. *Class* : kelas.

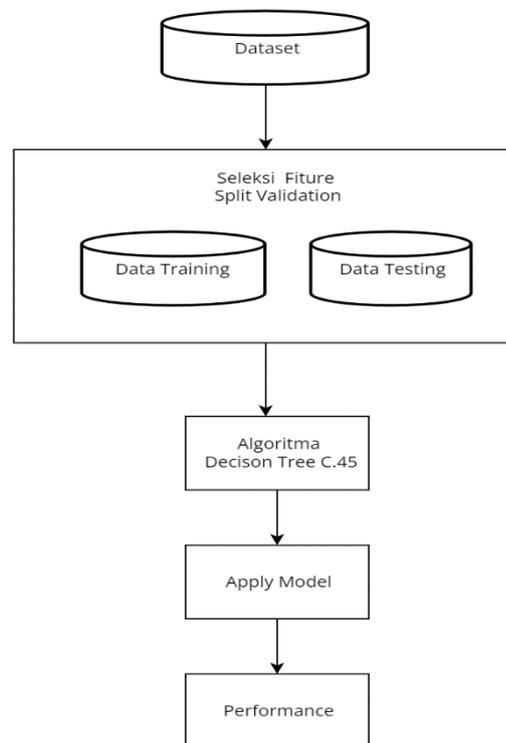
3. Pre-processing

Data Pre-processing adalah proses persiapan data. Proses ini sangat menentukan keberhasilan penambangan data. Pertama dilakukan pemilihan data, yaitu proses penghilangan atribut-atribut yang tidak digunakan. Selanjutnya dilakukan integrasi data, yaitu menggabungkan data yang akan diolah jika berasal dari beberapa database. Dalam penelitian ini, penulis tidak melakukan integrasi data karena data yang dipakai hanya berasal dari satu database. Setelah itu dilakukan pembersihan data, yaitu proses menghilangkan data yang tidak lengkap, data yang redundan, dan data yang kosong. Dalam penelitian ini penulis melakukan pembersihan data karena data yang diunduh sudah terdapat data yang kosong yaitu 16 record error dari 699 record menjadi 683 record, sehingga dataset yang diolah berjumlah 683 record. Setelah data dibersihkan, selanjutnya data ditransformasikan kedalam format yang sama. Pada penelitian ini, penulis mentransformasikan data menggunakan Microsoft Excel untuk membuat format data tabular

4. Proses Mining Menggunakan *Particle Swam Optimization (PSO)* dan algoritma Decision Tree C4.5 dengan menggunakan tools rapidminer.

Tahap selanjutnya menentukan metode dan pengujian metode yang akan digunakan untuk melakukan klasifikasi dengan teknik data mining untuk menemukan hasil akurasi terbaik.

Tahapan dalam melakukan uji coba pada rapidminer yang dilakukan peneliti adalah, memasukan dataset yang telah terseleksi kemudian seleksi fitur PSO, split data kemudian memasukan algoritmanya dalam hal ini menggunakan algoritma decision tree C.45 kemudian *apply model* dan terakhir *performance* tahapan uji coba tersebut dapat pula di lihat pada gambar bagan di bawah ini.



Gambar 3.2 Bagan Permodelan

5. *Interpretation and Evaluation* Setelah didapatkan rules, selanjutnya rules tersebut akan diinterpretasikan dan dievaluasi. Rules diinterpretasikan dengan cara dipreentasikan dengan bahasa yang mudah dipahami (Hidayati & Widi Nugroho, [s.d.]).

3.2. Alat dan Bahan

Penelitian ini menggunakan perangkat keras Laptop Swift SF314 – 56 G dengan processor intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz Installed RAM 8,00 GB (7,86 GB usable), sedangkan perangkat lunak yang digunakan Microsoft Excel dan Rapidminer Studio aplikasi yang digunakan untuk pengolahan data. Bahan yang digunakan adalah datase publik tentang kanker payudara yaitu di <https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28original%29>.