

## BAB III

### KERANGKA PENELITIAN

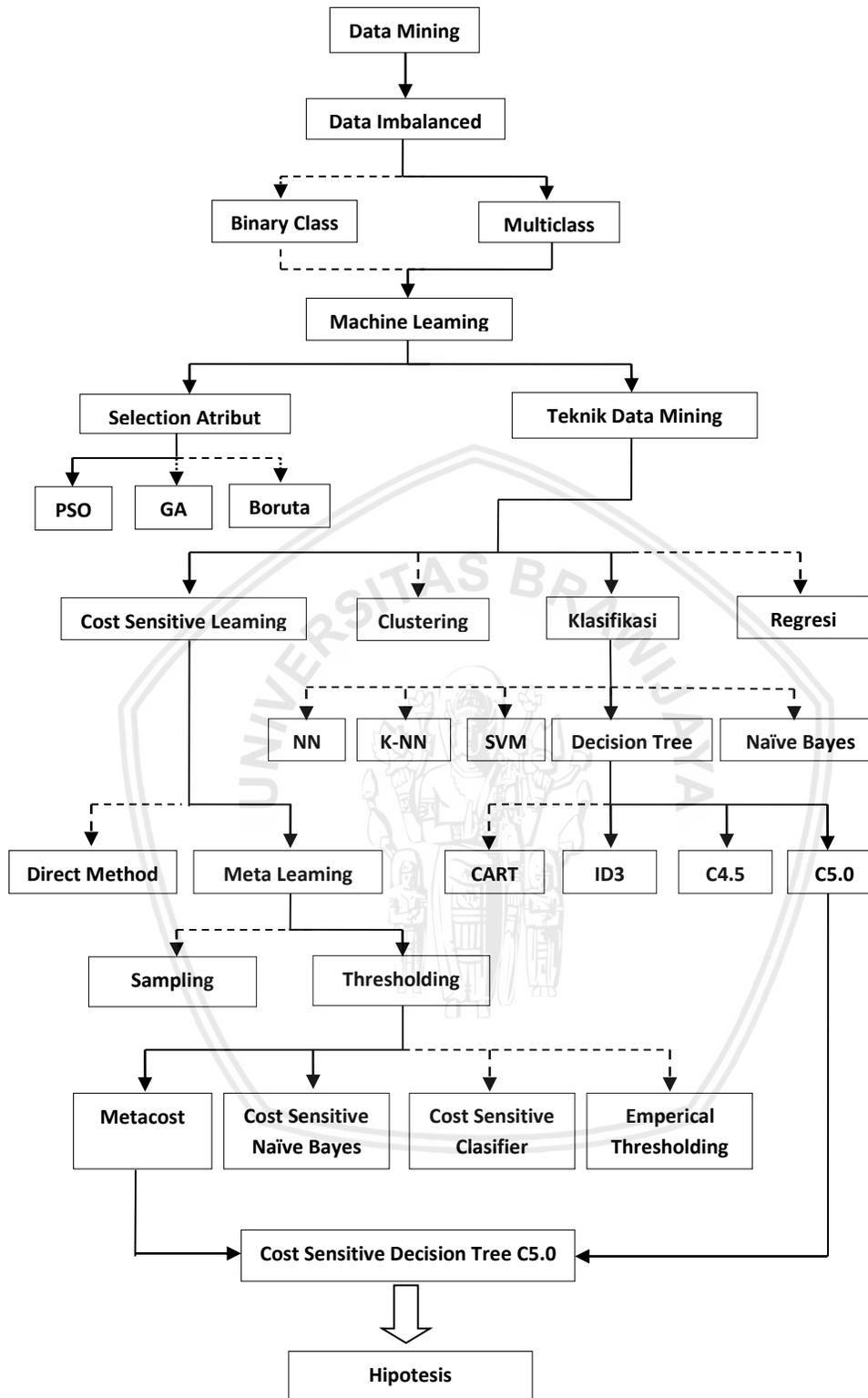
Berdasarkan latar belakang permasalahan dan landasan teori yang telah diuraikan dapat diperoleh suatu kerangka konsep penelitian sebagai landasan dasar berfikir dalam melakukan suatu penelitian yang dikembangkan dari tinjauan teori yang telah dibahas.

#### 3.1 Konsep Penelitian

Pemrosesan awal *dataset* dengan menghapus data *outlier* dan mengisi data *missing value*. Penghapusan data *outlier* dilakukan dengan menghapus data-data yang memiliki simpangan cukup jauh, namun ketika dilakukan pengujian, tidak terjadi perubahan *accuracy* yang signifikan. Pengisian *missing value* dilakukan dengan mengisi nilai rata-rata dari atribut tersebut.

Setelah dilakukan pemrosesan awal, kemudian dilakukan seleksi atribut. Seleksi atribut ini dilakukan menggunakan algoritma *particle swarm optimization* (PSO). *Dataset* yang telah diseleksi menggunakan PSO kemudian disiapkan untuk proses validasi data, dengan menggunakan metode *10-fold cross validation* dimana *dataset* hasil seleksi ini dibagi menjadi 90% data *training* dan 10% data *testing*.

Selanjutnya data *training* dilakukan pembelajaran untuk memperoleh pola model kemudian pola tersebut dilakukan pengujian menggunakan data *testing* dengan menggunakan metode *cost sensitive decision tree* dan *cost sensitive naïve bayes*. Hasil pengujian kemudian dievaluasi menggunakan parameter *accuracy*, *recall*, *precision*, *f-measure* dan *total cost*. Kerangka proses klasifikasi ditampilkan pada Gambar 3.1.



Gambar 3.1 Kerangka Konsep Penelitian

### 3.2 Hipotesis

Pada penelitian ini Berdasarkan permasalahan dan solusi metode yang telah dijelaskan dapat diperoleh kesimpulan sementara sebagai berikut:

- H1 : Pada proses seleksi atribut, *Particle swarm optimazation* (PSO) dapat menyeleksi atribut yang tidak relevan berdasarkan nilai bobot setiap atribut ( $w$ ) sehingga mampu meningkatkan performa *accuracy classifier*.
- H2 : Metode *decision tree* menggunakan algoritma C5.0 memiliki performa yang lebih baik untuk klasifikasi data *imbalanced multiclass* dibandingkan menggunakan algoritma C4.5 dan ID3.
- H3 : Metode *cost sensitive decision tree* C5.0 memiliki performa yang lebih baik untuk klasifikasi data *imbalanced multiclass* dibandingkan menggunakan algoritma *cost sensitive naïve bayes*.

