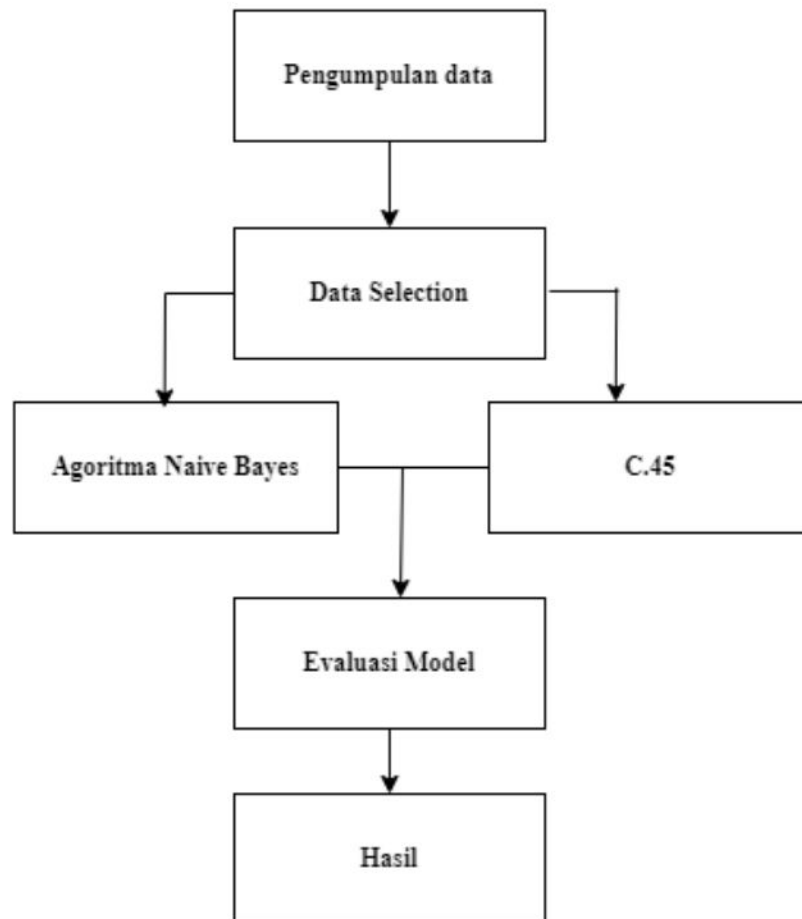


BAB III METODOLOGI PENELITIAN

3.1 Perancangan Penelitian

Adapun perancangan pada penelitian telah disusun sehingga alur dari proses penelitian ini sesuai dengan yang telah ditentukan.



Gambar 3. 1 Alur Penelitian[1]

Berikut penjelasan mengenai tahapan-tahapan yang ditunjukkan dalam diagram tersebut:

1. **Pengumpulan Data:** Tahapan pertama dalam proses ini adalah pengumpulan data. Data yang dikumpulkan bisa berasal dari berbagai sumber, seperti survei, observasi, atau data sekunder yang sudah ada. Proses ini penting karena data yang dikumpulkan akan digunakan dalam tahap-tahap selanjutnya.
2. **Data Selection:** Setelah data dikumpulkan, langkah berikutnya adalah pemilihan data (data selection). Pada tahap ini, data yang tidak relevan atau tidak memenuhi kriteria tertentu akan disaring atau dibuang, sementara data yang relevan akan dipilih untuk analisis lebih lanjut. Pemilihan data ini bertujuan untuk memastikan bahwa hanya data yang berguna yang akan digunakan dalam pemodelan.
3. **Algoritma Naïve Bayes :** Setelah Di Selection Langkah Selanjutnya adalah pengujian dengan Algoritma Naïve Bayes. Pada tahap ini, Langkah yang dilakukan pada tahap ini adalah menerjemahkan desain yang telah terbentuk menjadi sebuah system yang menerapkan algoritma Naïve Bayes.
4. **C4.5 :** Setelah Selection Langkah Selanjutnya juga harus dilakukan Pengujian dengan Algoritme C4.5. pada tahap ini Dilakukan pengujian untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak.
5. **Evaluasi Model:** Setelah data dipilih, langkah selanjutnya adalah evaluasi model. Di tahap ini, model analisis atau prediksi yang telah dikembangkan diuji atau dievaluasi menggunakan data yang telah dipilih. Evaluasi model bertujuan untuk menilai seberapa baik model bekerja dan apakah model tersebut memenuhi tujuan analisis.
6. **Hasil:** Tahapan terakhir adalah hasil, di mana output dari evaluasi model disajikan. Hasil ini bisa berupa laporan, visualisasi, atau interpretasi dari performa model dan implikasinya terhadap tujuan analisis.

3.2 Objek Penelitian

Objek penelitian merupakan suatu yang dijadikan fokus dalam melakukan suatu penelitian yang dijadikan objek atau fokus dalam penelitian ini yaitu mengenai pemilihan kriteria dengan mengimplementasikan algoritma *Naive Bayes* dan *C.45* dalam pengambilan keputusan seleksi penerima bantuan bedah rumah penduduk Desa.

3.3 Metode Pengumpulan data

Data Metode pengumpulan data adalah cara atau teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data. Untuk memperoleh data yang akurat dalam laporan penelitian ini, maka penulisan ini menggunakan beberapa metode yang dilakukan yaitu sebagai berikut:

- a. Observasi adalah suatu pengamatan sistematis yang dilakukan secara langsung terhadap objek di tempat penelitian. Dalam arti yang luas observasi dilakukan dengan cara mengamati langsung lokasi yang menjadi sumber informasi untuk memperoleh informasi yang dibutuhkan dalam penulisan ini. Beberapa informasi yang diperoleh dari hasil observasi adalah ruang (tempat), pelaku, kegiatan, objek, perbuatan, kejadian atau peristiwa waktu dan perasaan.
- b. Metode Wawancara Metode wawancara merupakan metode pengumpulan data dengan cara mengajukan beberapa pertanyaan langsung kepada narasumber. Wawancara dilakukan untuk mengetahui masalah yang timbul atau dialami langsung oleh yang bersangkutan. Dalam kegiatan ini diajukan pernyataan lisan dalam usaha untuk melengkapi data yang akan diperoleh. Wawancara dilakukan pada bagian-bagian yang terkait
- c. Metode Studi Pustaka Metode studi pustaka ini dilakukan dengan cara mempelajari jurnal, buku-buku atau studi literatur terdahulu yang berkaitan dengan, data mining C45 dan naive bayes classifier

Table 3 1. Pemilihan Variabel

| Variabel | Keterangan | Indikator | Detail Penggunaan |
|-----------------|---------------------|------------------|--------------------------|
| X1 | Jenis Dinding | V | Digunakan Atribut |
| X2 | Jenis Lantai | V | Digunakan Atribut |
| X3 | Bahan Bakar Memasak | V | Digunakan Atribut |
| X4 | Fasilitas MCK | V | Digunakan Atribut |
| X5 | Jumlah Tanggungan | V | Digunakan Atribut |
| X6 | Keterangan | V | Digunakan Labe |

Tabel diatas menjelaskan variabel yang akan digunakan dan tidak digunakan dalam penelitian ini. Indikator “V” adalah menandakan bahwa variabel tersebut akan digunakan.

1. Pre-Processing (data cleaning)

Tahap pre-processing merupakan tahap dilakukannya pembersihan terhadap data missing value yaitu data yang tidak konsisten atau kosong, dan juga dilakukannya pemilihan variabel terhadap data yang ingin digunakan pada proses data mining. Klasifikasi penerimaan program bantuan bedah rumah diatas akan menjadi keputusan layak renovasi atau tidak layak renovasi. Jumlah data yang digunakan adalah 589 data dengan 6 kriteria.

3.4 Naive Bayes

Langkah yang dilakukan pada tahapan ini adalah menerjemahkan desain yang telah terbentuk menjadi sebuah sistem yang menerapkan algoritma Naive Bayes.

Naïve Bayes Classifier adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema Bayes. Naïve Bayes merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik. Secara garis besar algoritma Naïve Bayes dapat dijelaskan seperti persamaan (1).

Rumus *Naïve Bayes* nya adalah:

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (2.2)$$

Keterangan:

X = data dengan kelas yang belum diketahui

H = hipotesis data X, merupakan suatu kelas yang spesifik

P(H|X) = probabilitas hipotesis H berdasar kondisi X (posteriori probability)

P(H) = probabilitas hipotesis H (posteriori probability)

P(X|H) = probabilitas X berdasar kondisi H

P(X) = probabilitas dari X

atau

$$Posterior Probability = \frac{Prior Probability \times likelihood}{evidence}$$

3.5 Algoritma C4.5

Algoritma C4.5 adalah sebuah algoritma yang berfungsi untuk membangun decision tree (pohon keputusan). Algoritma C4.5 dan pohon keputusan merupakan dua model yang tidak terpisahkan. Algoritma C4.5 adalah salah dari satu algoritma klasifikasi yang kuat dan banyak digunakan atau diimplementasikan untuk pengklasifikasian dalam berbagai hal. Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan (1996) sebagai versi perbaikan dari algoritma Iterative Dichotomiser 3 (ID3). Serangkaian perbaikan dilakukan pada algoritma ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan simple yang berpengaruh untuk pembentukan pohon keputusan. Perbaikan tersebut meliputi metode untuk

menangani data kontinu, mengatasi missing data, dan melakukan pemangkasan pohon.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i)$$

Keterangan:

S: Himpunan kasus

A: Data Atribut

n: Jumlah partisi di dalam atribut

|S_i|: Jumlah kasus pada partisi ke-i

|S|: Jumlah kasus

Sedangkan untuk menghitung nilai entropy dapat dihitung dengan rumus [16], persamaan 2.2.

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

Keterangan:

S: Himpunan kasus

n: Jumlah partisi dalam atribut

p_i: Proporsi dari S_i terhadap S

3.6 Pengujian Model *Decision Tree* C4.5 dan Pengujian Precision, Recall dan Accuracy Naive Bayes dan C.45

Pengujian metode dilakukan untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak. Pada penelitian ini proses pengujian dilakukan dengan 2 langkah yaitu, menggunakan perhitungan secara manual, kemudian data diuji menggunakan tools Rapid Miner untuk memastikan apakah hasil perhitungan manual dengan hasil yang diperoleh Rapid Miner sama atau tidak

Menghitung jumlah penerimaan bantuan bedah rumah dan tidak meneirma bantuan bedah rumah,serta entropy dari semua kasus dan semua dibagi berdasarkan atribut. Perhitungan nilai entropy total dapat dicari dengan rumus berikut.

3.7 Evaluasi Model

Pada tahap ini akan membahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Model yang terbentuk akan diuji dengan menggunakan Confusion Matrix untuk mengetahui tingkat akurasi. Confusion Matrix akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar dan prediksi negatif yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

3.8 Hasil

Hasil dari penelitian ini adalah pengujian yang dapat memperlihatkan nilai precision, recall dan accuracy dari algoritma Naive Bayes dan C.45.