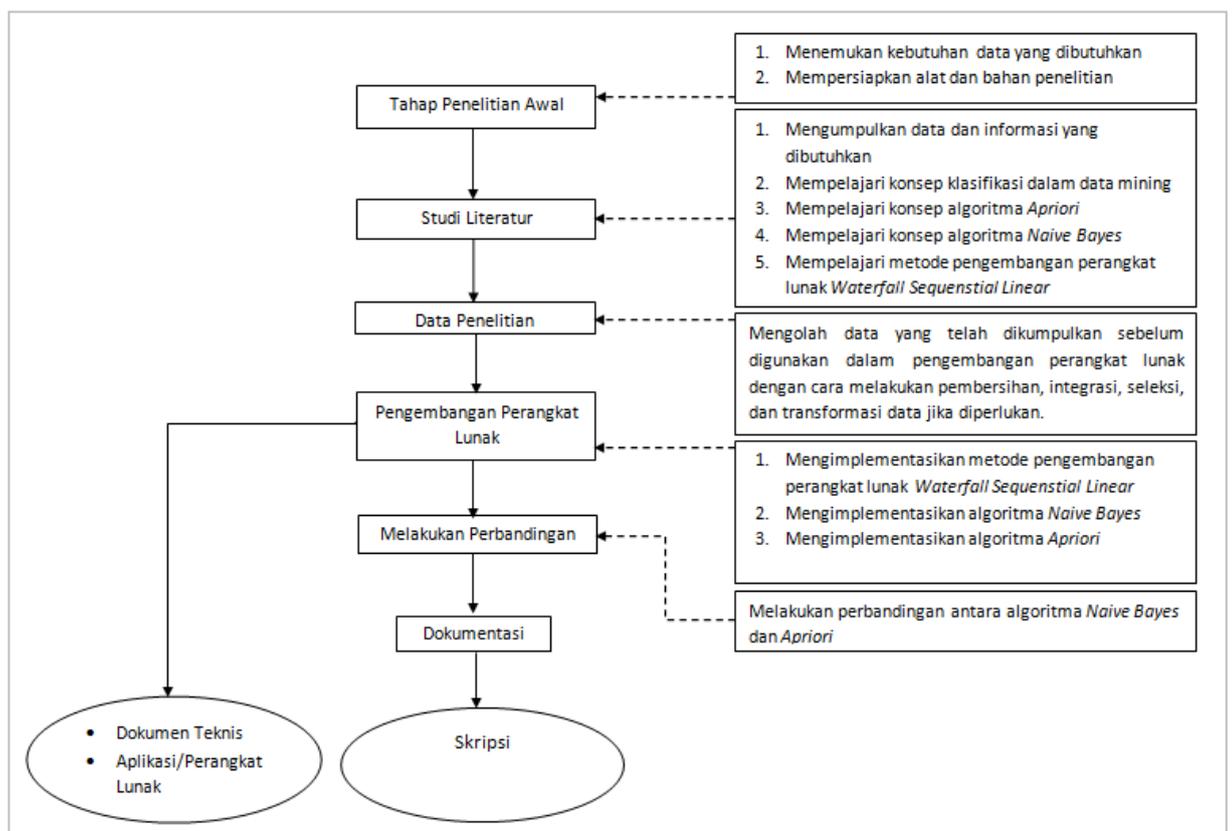


BAB III

METODOLOGI PENELITIAN

3.1 DESAIN PENELITIAN

Dalam melakukan penelitian, dibutuhkan desain penelitian agar penelitian yang dilakukan dapat berjalan dengan baik. Berikut ini merupakan desain penelitian yang digunakan pada proses rancang bangun aplikasi sistem pemilihan warna cat dinding kamar tidur berdasarkan kepribadian dengan menggunakan algoritma *Apriori* dan algoritma *Naive Bayes*.



Gambar 3.1 Desain Penelitian

Desain penelitian ini meliputi :

1. Tahapan Penelitian Awal
 - a. Menentukan data yang dibutuhkan

Data yang dibutuhkan adalah data demografi yang bersifat nominal atau setara. Data-data tersebut akan digunakan dalam proses klasifikasi *data mining*.

- b. Mempersiapkan alat dan bahan penelitian

Alat yang akan digunakan dalam penelitian ini adalah :

1) Laptop dengan spesifikasi:

- Prosesor Intel Core i3
- RAM 2,00 GB
- Harddisk 250 GB

2) Perangkat Lunak :

- Operating System : Windows 7 Professional
- Software : XAMPP, NetBeans

Bahan yang akan digunakan dalam penelitian ini adalah *paper*, *textbook*, dan dokumentasi lainnya yang didapat dari *World Wide Web*.

2. Studi Literatur

a. Mengumpulkan semua data dan informasi yang dibutuhkan

Metode pengumpulan data yang digunakan adalah *random sampling*. Pengumpulan data dilakukan dengan cara menyebar survey secara acak hingga jumlah sampel yang dibutuhkan terpenuhi. Survey adalah kegiatan pengumpulan data dari sebagian populasi (penduduk) yang pemilihannya dilakukan dengan menggunakan metode statistik tertentu sehingga tetap dapat melakukan pendugaan atas populasinya. Untuk menentukan jumlah sampel yang dibutuhkan, digunakan rumus *cross sectional* sebagai berikut:

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} p (1-p) N}{d^2(N-1) + Z^2_{1-\frac{\alpha}{2}} p (1-p)}$$

Namun, jika besar populasi (N) tidak diketahui, maka besar sampel dihitung dengan rumus :

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} p q}{d^2} = \frac{Z^2_{1-\frac{\alpha}{2}} p (1-p)}{d^2} \quad \begin{matrix} \text{(Snedecor GW \& Cochran WG, 1967)} \\ \text{(Lemeshowb dkk, 1997)} \end{matrix}$$

Keterangan :

n = jumlah sampel minimal yang diperlukan

Z = score Z, berdasarkan nilai α yang diinginkan

α = derajat kepercayaan

d = toleransi kesalahan

p = proporsi kasus yang diteliti dalam populasi, jika p tidak diketahui maka gunakan p terbesar. p terbesar yaitu $p = 0.5$

$1-p = q$, yaitu proporsi untuk terjadinya suatu kejadian. Jika penelitian ini menggunakan p terbesar, maka $q = 1-p = 1-0.5$

Batas toleransi kesalahan dinyatakan dengan prosentase. Semakin kecil toleransi kesalahan, semakin akurat sampel menggambarkan populasi. Misalnya penelitian dengan batas kesalahan 5% berarti memiliki tingkat akurasi 95%. Pada penelitian ini, ditentukan bahwa batas toleransi kesalahan adalah $5\% = 0.05$.

Besar nilai Z disesuaikan dengan nilai α :

Tabel 3.1 Besar nilai Z disesuaikan dengan Nilai α

α	$1 - \alpha$	$Z_{1 - \alpha/2}$	$Z_{1 - \alpha}$
1%	99%	2.58	2.33
5%	95%	1.96	1.64
10%	90%	1.64	1.28

Sesuai dengan rumus *cross sectional* dimana $Z_{1 - \alpha/2}$, maka besaran score Z yang akan diambil adalah sesuai dengan kolom ketiga. Pada penelitian ini, derajat kepercayaan yang digunakan adalah 5%, maka $Z_{1 - \alpha/2} = 1.96$ Jika sudah ditetapkan bawah score $Z = 1.96$, maka $Z^2 = 3.84$ atau dibulatkan menjadi 4. Maka, rumus untuk mengetahui jumlah sampel yang dibutuhkan sering kali dtuliskan sebagai berikut:

$$n = \frac{4 p q}{d^2}$$

Maka, dari rumus di atas jumlah sampel yang dibutuhkan pada penelitian ini adalah:

$$\begin{aligned} n &= \frac{4 \cdot 0.5 (1-0.5)}{0.05^2} \\ n &= \frac{0.9604}{0.0025} \\ n &= 384 \end{aligned}$$

Jadi, jumlah sampel minimal yang dibutuhkan untuk penelitian ini berjumlah 384 sampel.

Sampel penelitian tersebut akan dijadikan sebagai objek penelitian pada penelitian ini. Sampel penelitian tersebut terdiri dari 4 atribut berjenis data nominal, yaitu dari jenis kelamin, umur, jenis psikologi manusia, dan golongan warna.

b. Mempelajari konsep klasifikasi dalam *data mining*.

Inti konsep dari *data mining* adalah menggali dan mendapatkan pengetahuan (*knowledge*) dari kumpulan data-data yang ada. *Data mining* dapat dikategorikan sebagai suatu proses pendukung pengambil keputusan dimana kita mencari pola informasi dalam data. Pencarian tersebut disebut sebagai *discovery*. *Discovery*

adalah proses pencarian dalam basis data untuk menemukan pola yang tersembunyi. *Discovery* dapat dilakukan dengan cara klasifikasi. Pada proses klasifikasi, *record-record* data yang sebelumnya tidak terlihat dinyatakan kelasnya seakuran mungkin. Pada penelitian skripsi ini, klasifikasi digunakan dalam pemodelan prediktif sehingga dapat memprediksi label kelas untuk *record* yang belum diketahui kelasnya.

c. Mempelajari konsep algoritma *Naive Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive*, dimana *Naive* itu sendiri diasumsikan sebagai kondisi antar atribut saling bebas. Pada proses algoritma *Naive Bayes*, sejumlah petunjuk yang disebut sebagai atribut diperlukan untuk membantu dalam membentuk kelas yang cocok bagi sampel yang dianalisis.

d. Mempelajari konsep algoritma *Apriori*

Inti konsep dari algoritma *Apriori* adalah melakukan pencarian *frequent itemset* dengan menggunakan teknik *association rule* (Erwin, 2009). Algoritma ini ditujukan untuk mencari kombinasi *itemset* yang mempunyai suatu nilai keseringan tertentu sesuai kriteria diinginkan. Algoritma *Apriori* menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses

informasi selanjutnya. Pada algoritma *Apriori* menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum *support* dan minimum *confidence*. *Support* adalah nilai pengunjung atau persentase kombinasi sebuah *item* dalam *database*. Sedangkan *confidence* adalah nilai kepastian yaitu kuatnya hubungan antar *item* dalam sebuah *Apriori*.

- e. Mempelajari konsep metode pengembangan perangkat lunak Metode *Waterfall Sequential Linear*

Model metode rekayasa perangkat lunak (RPL) ini sering disebut dengan “*classic life cycle*” atau model *waterfall*. Karena model ini diperkenalkan pertama kali pada tahun 70-an maka sering dianggap kuno, tetapi merupakan model yang paling banyak dipakai didalam *Software Engineering (SE)*. Model ini melakukan pendekatan secara sistematis dan urut mulai dari level kebutuhan sistem lalu menuju ke tahap analisis, desain, *coding*, *testing/verification*. Disebut dengan *waterfall* karena tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya dan berjalan berurutan. Sebagai contoh tahap desain harus menunggu selesainya tahap sebelumnya yaitu tahap *requirement* yang terdiri dari tahap-tahap: analisis (*analysis*), desain (*design*), kode (*code*), dan uji coba (*test*).

3. Data Penelitian

Setelah dilakukan survey untuk mendapatkan jumlah minimal sampel yang dibutuhkan untuk penelitian, data-data tersebut kemudian diolah sebelum

digunakan dalam pengembangan perangkat lunak dengan cara melakukan pembersihan, integrasi, seleksi, dan transformasi data jika diperlukan.

Jika terdapat data yang bersifat interval, maka data tersebut dibuatkan tabel distribusi frekuensi. Fungsi dari tabel distribusi frekuensi adalah agar data yang bersifat interval dapat disajikan dengan baik sehingga dapat dengan mudah dipahami serta dapat memberikan potret yang lebih jelas berkaitan dengan distribusi data. Untuk membuat tabel distribusi frekuensi dapat dilakukan langkah-langkah berikut ini:

- a) Urutkan data, biasanya data diurutkan dari nilai yang paling kecil. Tujuan dari pengurutan data ini adalah agar *range* data dapat diketahui dan mempermudah penghitungan frekuensi tiap kelas.
- b) Tentukan *range* (rentang atau jangkauan). Menentukan *range* dengan cara mencari selisih antara data terendah dan data tertinggi.
- c) Tentukan banyak kelas interval yang diinginkan. Dalam menentukan kelas janganlah terlalu banyak atau terlalu sedikit. Tentukan kelas antara 5 sampai dengan 20 kelas, tergantung banyak dan sebaran data. Untuk menentukan banyaknya kelas dapat menggunakan aturan Struges:

$$\text{Banyaknya kelas} = 1 + 3,3 \log n$$

- d) Tentukan panjang atau lebar kelas interval (P):

$$P = R / n$$

Keterangan :

P = Panjang kelas interval

R = Rentang (*range*)

n = Banyak kelas interval

- e) Tentukan batas bawah kelas interval pertama.

Tidak ada aturan yang mengikat untuk menentukan nilai batas bawah kelas interval pertama, asalkan nilai terkecil masih masuk ke dalam kelas tersebut.

Maka, dari langkah-langkah di atas dapat diperoleh informasi berupa banyaknya kelas, panjang kelas, dan batas bawah kelas. Dengan informasi-informasi tersebut, maka tabel distribusi frekuensi dapat dibuat.

4. Pengembangan Perangkat Lunak

Setelah tahapan diatas dijalankan, tahapan selanjutnya yaitu pengembangan perangkat lunak. Dalam mengembangkan sebuah perangkat lunak, dibutuhkan sebuah metode agar pengembangan perangkat lunak dapat berjalan dengan baik. Metode tersebut terdiri dari proses-proses terstruktur yang meliputi : Analisis,

Desain, Implementasi, Pemeliharaan, dan Pengujian (*Analysis, Design, Code, dan Testing*). Proses-proses terstruktur tersebut dapat dijabarkan sebagai berikut.

a. Analisis

Dalam tahap ini dilakukan analisis pembangunan sistem terhadap kebutuhan pengguna, dan kebutuhan perangkat lunak yang meliputi penganalisaan informasi, unjuk kerja, dan antar muka (*interface*) yang diperlukan.

b. Desain

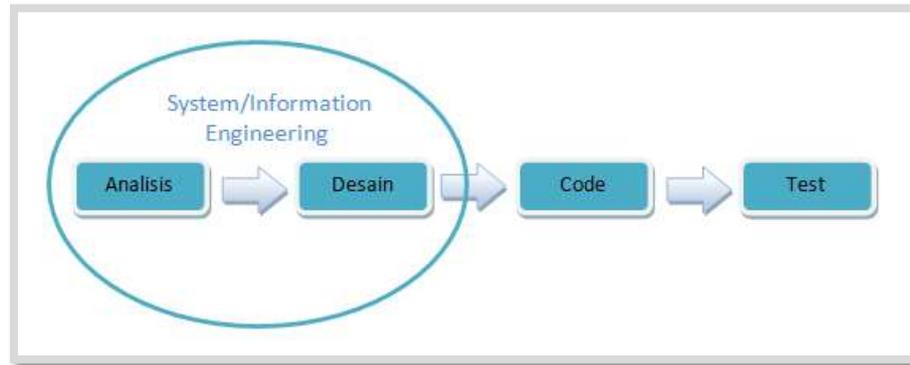
Pada tahap ini dilakukan perancangan struktur data, arsitektur perangkat lunak, representasi antarmuka, dan algoritma prosedural pada sistem.

c. Code

Tahap ini merupakan fase menterjemahkan model atau desain yang telah ditetapkan kedalam bahasa yang dimengerti komputer, dalam penelitian ini menggunakan PHP dan MySQL.

d. Test

Pada tahap ini dilakukan pengujian yang difokuskan pada logika internal perangkat lunak yaitu memastikan bahwa semua pernyataan sudah diuji. Sedangkan pada fungsi eksternal yaitu mengarahkan pengujian untuk menemukan kesalahan-kesalahan dan memastikan bahwa masukan yang dibatasi akan memberikan hasil aktual yang sesuai dengan hasil yang dibutuhkan.



Gambar 3.2 Metode *Waterfall Sequential Linear*

Selain itu, pada tahap tes ini juga dilakukan pengujian terhadap tingkat akurasi pada masing-masing algoritma yang diterapkan pada perangkat lunak ini. Untuk melakukan pengukuran tingkat akurasi, akan diberikan 50 data set yang akan diprediksi kelasnya. Dari ke-50 data tersebut akan dihitung seberapa besar prosentase algoritma tersebut merekomendasikan kelas dengan benar.

Dengan cara demikian, maka akan diketahui algoritma mana yang memiliki tingkat akurasi yang lebih tinggi dan *error rate* yang lebih rendah serta diketahui pula algoritma mana yang memiliki tingkat akurasi lebih baik.