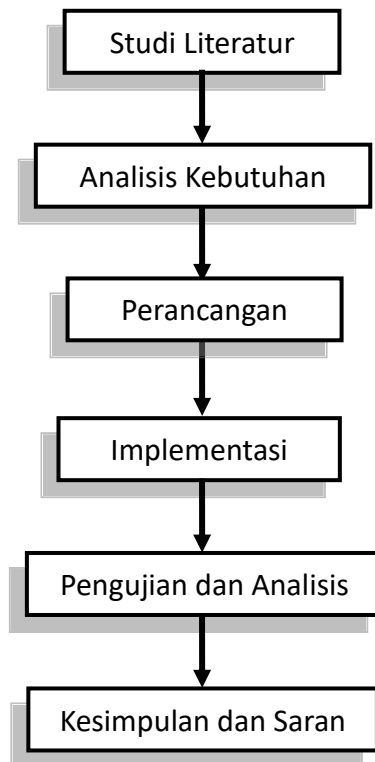


## BAB 3 METODOLOGI

Dalam bab ini membahas langkah-langkah pengerjaan penelitian penentuan jenis kanker berdasarkan struktur protein menggunakan algoritma *Naive Bayes*. Berikut adalah diagram alir dari penelitian yang akan dilakukan.



**Gambar 3.1 Diagram Alir Metodologi Penelitian**

### 3.1 Studi Literatur

Studi literatur bertujuan mendapatkan pengetahuan tentang konsep dasar bahasan dari penelitian sebagai sumber penulisan, pemecahan masalah dan pemahaman skripsi. Referensi dapat berupa buku, penelitian, dan jurnal ilmiah. Topik dan teori yang mendukung penulisan skripsi ini meliputi:

1. Kanker
2. Protein tubuh
3. Gen P53
4. Data Mining
5. *Naive Bayes*

### 3.2 Analisis Kebutuhan

Analisis kebutuhan dilakukan untuk mengetahui kebutuhan dalam membuat aplikasi serta untuk mengetahui kebutuhan pengguna. Salah satunya adalah pengumpulan data. Proses pengumpulan data dilakukan untuk mendapatkan data yang sesuai dengan tujuan dari penelitian. Pada penelitian kali ini akan dibuat aplikasi yaitu tentang prediksi atau klasifikasi jenis kanker berdasarkan struktur protein dari pasien. Data yang dibutuhkan berupa data latih dan data uji pasien kanker dan non-kanker. Kemudian data yang sudah didapatkan akan diproses menggunakan algoritma *Naive Bayes*.

### 3.3 Perancangan

Perancangan sistem dilakukan untuk menjelaskan alur proses pembangunan sistem secara rinci. Sistem yang akan dibuat adalah perangkat lunak untuk melakukan klasifikasi terhadap susunan protein menggunakan algoritma *Naive Bayes*. Diperlukan sistem yang dapat memproses masukan berupa data latih dan data uji dan menghasilkan keluaran berupa jenis kanker dari pasien. Untuk memenuhi kebutuhan dari sistem, perancangan sistem yang diperlukan adalah perancangan algoritma, perancangan antarmuka dan perancangan pengujian. Perancangan algoritma *Naive Bayes* meliputi pencarian *likelihood* dan *prior*. Perancangan antarmuka dilakukan untuk mempermudah pengguna dalam memahami jalannya sistem. Pada perancangan pengujian akan dijelaskan bagaimana jalannya pengujian pada sistem yang sudah selesai.

### 3.4 Implementasi

Semua yang telah diimplementasikan pada tahap ini berdasarkan pada perancangan sistem. Sistem dibangun menggunakan bahasa pemrograman Java dengan menggunakan bantuan perangkat lunak Netbeans IDE 8.1. Pada implementasi akan dibuat *user interface* sistem *desktop*. Pengolahan dataset protein termasuk membaca data (data latih dan data uji) dari file *excel*. Kemudian dilakukan implementasi perhitungan-perhitungan *Naive Bayes* antara lain perhitungan probabilitas awal (*likelihood*), probabilitas kelas (*prior*) dan probabilitas akhir (*posterior*). Juga diimplementasikan perhitungan akurasi dari sistem yang membandingkan hasil klasifikasi dan kelas sebenarnya.

### 3.5 Pengujian dan Analisis

Jenis pengujian yang digunakan untuk menguji sistem adalah pengujian validasi dan pengujian akurasi. Pengujian validasi dibutuhkan oleh sistem untuk mengetahui apakah fungsi-fungsi yang ada pada sistem berjalan dengan baik dan sesuai dengan fungsinya. Jika pada saat pengujian tidak ditemukan error, berarti sistem tersebut sudah baik. Pada pengujian tingkat akurasi akan dilakukan pengujian tingkat akurasi terhadap jumlah data uji dan terhadap jumlah dataset. Terdapat 5 dataset, yaitu 320, 400, 480, 588 dan 848 data. Dari

masing-masing dataset data uji yang digunakan disesuaikan yaitu 10% hingga 60%. Kemudian setelah hasil pengujian keluar, akan dilakukan analisis pada masing-masing data.

### **3.6 Kesimpulan dan Saran**

Kesimpulan dan saran merupakan bagian akhir dari penelitian yang merupakan jawaban atas masalah yang telah diutarakan pada latar belakang dan penemuan kekurangan pada saat pengerjaan penelitian agar jika akan menggunakan data atau algoritma yang sama dapat menjadi bahan masukan sehingga kualitas penelitian selanjutnya menjadi lebih baik.