

# BAB 1 PENDAHULUAN

## 1.1 Latar belakang

Kanker merupakan penyakit yang kerap menjadi momok bagi sebagian besar orang memang telah memakan banyak korban. Tidak mengenal dari kalangan bawah ataupun kalangan pejabat hingga artis. Semakin berkembangnya zaman semakin berkembang pula bermacam-macam penyakit. Hal ini dikarenakan semakin banyak virus yang tersebar di masyarakat. Virus ini tersebar dengan berbagai cara. Mulai dari gaya hidup, keturunan, atau bawaan hormon dalam tubuh.

Kanker adalah istilah yang digunakan untuk menggambarkan ratusan penyakit berbeda dengan fitur tertentu yang sama. Kanker dimulai dengan perubahan dalam struktur dan fungsi sel yang menyebabkan sel membelah dan menggandakan diri tanpa terkontrol. Sel kemudian dapat menyerang dan merusak jaringan sekitar, dan sel dapat memisahkan diri dan menyebar ke area lain dalam tubuh. Umumnya kanker dinamai sesuai organ dan jenisnya tempat pertama kali ia berkembang (Kelvin, Joanne Frankel., dkk., 2011).

Saat ini semakin banyak perkumpulan yang mengatasnamakan diri mereka sebagai *cancer survivor*. Perkumpulan atau organisasi ini terdiri dari orang-orang yang pernah atau sedang menderita penyakit kanker, tetapi mereka melakukan berbagai macam pengobatan dengan tujuan untuk bertahan hidup demi melawan penyakit kanker. Anggota dalam organisasi ini biasanya berbagi cerita dan pengalaman agar orang yang baru terkena penyakit kanker tidak merasa terpuruk atau putus asa. Mereka memberikan keyakinan bahwa penyakit yang terkenal ganas dan mematikan ini sebenarnya bisa disembuhkan dengan gaya hidup sehat.

Telah diketahui bahwa sekumpulan faktor genetik dan lingkungan dapat meningkatkan risiko terjadinya kanker. Kode genetik yang mengatur pembentukan sel-sel kanker berhasil dipecahkan. Meski bentuk pengobatan dengan target gen masih terus dikembangkan, temuan ini membawa harapan baru untuk melawan penyakit mematikan tersebut. Peta genetik yang dibuat para ilmuwan Amerika Serikat menunjukkan setidaknya terdapat sekitar 200 gen termutasi yang berperan mengatur pembentukan, pertumbuhan, dan penyebaran tumor. Sebagian besar gen yang termutasi ini belum diketahui sebelumnya (Diananda, Rama, 2011).

Mutasi gen yang paling sering ditemukan pada kanker manusia adalah Gen P53. Gen P53 merupakan gen penekan tumor yang mengkode atau mengekspresikan protein 53, nama ini diambil dari berat molekulnya yang sebesar 53 kilodalton yang dikode oleh gen p53 yakni gen pengontrol stabilitas sel. Protein p53 merupakan faktor transkripsi terhadap gen-gen yang terlibat dalam regulasi siklus sel, induksi apoptosis, repair DNA, dan stabilitas genome. Mutasi gen p53 merupakan abnormalitas molekuler tersering pada lebih dari 50% kasus keganasan. Hilangnya fungsi p53 akibat dari mutasi dapat menimbulkan transformasi keganasan, metastase tumor dan resistensi terhadap kemoterapi.

Sebagian besar penderita menyadari atau merasakan bahwa dirinya terjangkit kanker adalah jika gen kanker tersebut telah berkembang luas dan mulai mengganggu organ tubuh yang lain. Dampaknya adalah banyak jiwa yang terlambat untuk diselamatkan. Untuk menghindari dan mengurangi tingkat kematian, pemeriksaan dini seharusnya wajib dilakukan oleh setiap orang agar jika terdapat gejala-gejala yang mengarah pada kanker dapat segera diberi tindakan pencegahan dan penyembuhan.

Dengan semakin berkembang pesatnya kecanggihan teknologi juga diharapkan adanya sebuah sistem atau aplikasi yang mampu mengidentifikasi penyakit kanker. Pengidentifikasi dilakukan dengan cara mengclusterkan data pasien ke dalam kelas "*non-cancer*", "*breast cancer*", "*colorectal cancer*" dan kelas "*lung cancer*". Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas pada objek-objek yang memiliki kemiripan. Data dalam satu kelas memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan data yang ada pada kelas yang lain.

Dari hasil pengklusteran data uji tersebut maka akan didapatkan hasil termasuk ke dalam kelas manakah pasien tersebut. Dengan syarat bahwa semua dataset memiliki panjang data yang sama sehingga pengklusteran dapat dilakukan dengan baik.

Dalam melakukan pengklusteran juga tidak bisa dilakukan dengan sembarang cara. Banyak terdapat metode-metode dan perhitungan algoritma yang dapat mendukung proses pengklusteran data tersebut. Metode yang sering digunakan untuk pengklusteran adalah algoritma K-Means. Algoritma K-Means adalah algoritma yang membagi data secara otomatis ke dalam sejumlah K kelas (Wibisono, Yudi, 2011). Tetapi algoritma ini memiliki beberapa kelemahan sehingga terdapat beberapa algoritma ataupun metode sebagai penyempurna K-Means, salah satunya adalah metode K-Medoids.

K-Medoids adalah sebuah algoritma yang menggunakan metode partisi clustering untuk mengelompokkan sejumlah n objek menjadi k cluster. Metode ini telah digunakan pada penelitian sebelumnya untuk pengklasifikasian penyakit kanker berdasarkan kandungan logam dalam darah. Dan tingkat akurasi mencapai 57.14% (Nastiti, Shofi., dkk., 2014). Sedangkan dengan data kanker yang sama yaitu berdasarkan struktur protein dalam tubuh tetapi menggunakan algoritma Modified K-Nearest Neighbor (MKNN) tingkat akurasi maksimum yang dicapai adalah 43.53% (Retwitasari, Arintha, 2016).

Berdasarkan latar belakang yang telah dipaparkan terdapat beberapa masalah yang bisa disimpulkan, dan juga berdasarkan penelitian sebelumnya yang telah dilakukan dengan melihat dari tingkat akurasi maka judul yang bisa diambil adalah "**Clustering Pasien Kanker Berdasarkan Struktur Protein Dalam Tubuh Menggunakan Metode K-Medoids**".

## 1.2 Rumusan masalah

1. Bagaimana mengimplementasikan metode K-Medoids untuk clusterisasi pasien kanker berdasarkan struktur protein dalam tubuh?
2. Bagaimana kualitas cluster yang dihasilkan oleh metode K-Medoids untuk clusterisasi pasien kanker berdasarkan struktur protein dalam tubuh?

## 1.3 Tujuan

1. Mengimplementasikan metode K-Medoids untuk clustering data pasien kanker berdasarkan struktur protein dalam tubuh.
2. Mengetahui hasil uji kualitas cluster yang dihasilkan oleh sistem dalam melakukan clusterisasi pasien kanker berdasarkan struktur protein dalam tubuh menggunakan metode K-Medoids.

## 1.4 Manfaat

Manfaat yang dapat diambil dari penelitian ini adalah memudahkan dalam pengclusteran pasien kanker dalam kelas *non-cancer*, *breast cancer*, *colorectal cancer*, atau *lung cancer* berdasarkan struktur protein dalam tubuh dengan menggunakan metode K-Medoids.

## 1.5 Batasan masalah

Pada penelitian ini diberikan batasan masalah agar penelitian tidak semakin meluas dan tidak terarah. Batasan masalah tersebut mencakup :

1. Metode yang diimplementasikan untuk clusterisasi pasien kanker adalah metode K-Medoids.
2. Clusterisasi pasien kanker ini berdasarkan pada struktur protein dalam tubuh.
3. Hanya digunakan 4 cluster yaitu "*non-cancer*", "*breast cancer*", "*colorectal cancer*", dan "*lung cancer*".
4. Dilakukan sebuah pengujian *Silhouette Coefficient* untuk mengetahui kualitas cluster yang dihasilkan sistem.
5. Semua data dari susunan protein yang digunakan pada data uji mempunyai panjang yang sama.

## 1.6 Sistematika pembahasan

### BAB 1 PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika pembahasan.

### BAB 2 LANDASAN KEPUSTAKAAN

Landasan kepustakaan berisi uraian dan pembahasan tentang teori, konsep, model, metode, atau sistem dari literatur ilmiah, yang berkaitan dengan tema, masalah, atau pertanyaan penelitian yaitu tentang Clustering Pasien Kanker Berdasarkan Struktur Protein Dalam Tubuh Menggunakan Metode K-Medoids.

### **BAB 3 METODOLOGI**

Berisi langkah-langkah, metode, dan teknik yang digunakan dalam menyelesaikan masalah yang terjadi beserta penjelasannya.

### **BAB 4 PERANCANGAN**

Dalam implementasi pengembangan dibuat rancangan arsitektur terlebih dahulu. Mulai dari analisis kebutuhan yang berisi deskripsi sistem, analisis kebutuhan data, dan identifikasi aktor. Juga dijelaskan tentang perancangan perangkat lunak yang dilakukan yaitu perancangan algoritma, perhitungan manual dengan metode K-Medoids, dan perancangan antarmuka. Yang terakhir akan dijelaskan tentang pengujian perangkat lunak yang dilakukan. Terdapat dua jenis pengujian yaitu pengujian pengaruh jumlah data dan pengujian kualitas cluster.

### **BAB 5 IMPLEMENTASI**

Dalam bab implementasi akan dijelaskan dan ditampilkan hasil dari proses perancangan yang telah dilakukan. Juga akan dijelaskan sistem kerja metode yang digunakan dalam hal ini adalah metode K-Medoids. Setelah sistem jadi dan bisa dioperasikan maka akan dilakukan pengujian kualitas clustering yang dihasilkan oleh metode K-Medoids.

### **BAB 6 PENGUJIAN**

Setelah sistem telah selesai dibuat maka akan dilakukan pengujian sistem dan evaluasi sebelum sistem benar-benar siap dirilis. Terdapat dua pengujian yang dilakukan, yang pertama adalah pengujian kualitas clustering yang dilakukan menggunakan *Silhouette Coefficient* dan pengujian pengaruh jumlah data latih.

### **BAB 7 PENUTUP**

Bagian penutup berisi kesimpulan dan saran tentang hasil penelitian yang diperoleh berdasarkan tujuannya.