



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Menurut *The American Cancer Society* [1], kanker merupakan terjadinya pertumbuhan sel dalam tubuh secara terus menerus tanpa adanya pergantian terhadap sel lama atau sel yang telah rusak. Penyakit leukemia merupakan salah satu jenis kanker yang terjadi pada sel darah manusia [2]. Berdasarkan data *World Health Organization* (WHO), penderita penyakit leukemia di negara industri memperoleh angka 13,5% dari total penderita kanker pada anak-anak usia di bawah 15 tahun. Sedangkan di negara berkembang, termasuk negara Indonesia, penderita penyakit leukemia memperoleh angka 15,7% pada anak-anak [3]. Selain itu, menurut *The American Cancer Society* [4], terdapat sekitar 59,610 kasus baru penyakit leukemia (seluruh jenis leukemia) dan 23,710 kematian yang terjadi akibat dari penyakit leukemia di Amerika. Sehingga dapat terlihat bahwa pada anak-anak penderita penyakit kanker, leukemia memiliki persentase yang cukup tinggi sehingga perlu dilakukan pencarian obat yang dapat mengatasi penyakit tersebut dalam waktu yang singkat.

Suatu penyakit kanker timbul dikarenakan terjadi mutasi terhadap protein dalam tubuh. Protein yang mengalami mutasi dalam tubuh tersebut disebut sebagai *biomarker*, yaitu indikasi secara objektif oleh pengamatan keadaan medis seorang pasien yang diukur secara akurat dan berulang [5]. Dengan dilakukan penelitian untuk menemukan protein signifikan yang memengaruhi suatu penyakit maka dapat ditemukan *biomarker* terhadap penyakit tersebut [6]. Protein signifikan merupakan protein-protein yang memiliki peranan penting dalam suatu jaringan interaksi antar protein (PPI) [7]. Untuk mendeteksi PPI yang terjadi seringkali sulit untuk dilakukan oleh para peneliti karena cenderung membutuhkan waktu yang lama dan juga biaya yang besar untuk mendukung keberhasilan penelitian [8]. Oleh karena itu, dibutuhkan alternatif lain dalam melakukan penelitian identifikasi protein signifikan terhadap suatu penyakit dengan cepat dan cenderung murah yaitu dengan metode pendekatan komputasi berupa teknik kecerdasan buatan atau *artificial intelligence* (AI) [9][10].

Bioinformatika adalah suatu pendekatan komputasi yang digunakan dalam melakukan pengumpulan dan analisa data biologi yang bersifat biologis seperti

DNA, RNA, dan interaksi antar protein [11]. Pendekatan komputasi mencakup dalam penerapan metode di bidang matematika, statistika, dan informatika untuk memecahkan masalah-masalah biologi yang sulit untuk dilakukan secara manual [12]. Bioinformatika sudah mulai masuk dan berkembang di Indonesia, terutama di kalangan mahasiswa yang sudah mulai banyak melakukan penelitian terhadap bidang ini. Seperti penelitian yang telah dilakukan sebelumnya [7][13][14] mengenai identifikasi protein signifikan dengan menggunakan pembelajaran mesin *Principal Component Analysis* (PCA) dan analisis topologi interaksi protein sebagai teknik pendekatan. Ada pula penelitian terdahulu yang telah dilakukan dengan menggunakan pembelajaran mesin untuk mengidentifikasi protein signifikan, *clustering* protein dari suatu penyakit, seleksi fitur, hingga menemukan biomarker dari suatu penyakit seperti [10][15][16][17][18][19] dengan menggunakan PCA, [15][20][18] dengan menggunakan *Support Vector Machine* (SVM), [21][22] dengan menggunakan *Random Forest*, [23][24] dengan menggunakan *Hierarchical Cluster Analysis* (HCA), dan [25][26] dengan menggunakan *ClusterONE*.

Pada penelitian ini, untuk melakukan identifikasi protein signifikan terhadap penyakit leukemia digunakan metode PCA dan *ClusterONE* sebagai pendekatan *clustering* karena dianggap sebagai metode yang sesuai dengan kebutuhan dataset yang didapatkan. Sebelum menggunakan metode PCA, dilakukan pembentukan graf PPI dengan menggunakan perhitungan *centrality measurements*. Metode PCA dipilih sebagai metode yang akan digunakan dikarenakan dapat mereduksi fitur dengan cepat dan tanpa menghilangkan karakteristik atau informasi penting dalam fitur [27], namun pada penelitian ini metode PCA digunakan untuk memberikan bobot kepada setiap nilai *centrality measurements*. Selain itu, untuk pendekatan *clustering* terhadap graf PPI yang telah terbentuk digunakan algoritma *ClusterONE* dikarenakan pendekatan ini dianggap dapat memberikan hasil akurasi yang lebih tinggi dan memberikan *one-to-one mapping* yang lebih baik terhadap setiap *cluster* data [28]. Melalui penelitian ini, diharapkan metode PCA dan *ClusterONE* dapat memberikan hasil yang terbaik dalam melakukan identifikasi protein signifikan penyakit leukemia.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan sebelumnya, berikut rumusan masalah pada penelitian ini.

1. Bagaimana membentuk subgraf PPI berdasarkan *Centrality Measurements*

dan melakukan *Clustering* berdasarkan pada protein yang berhubungan dengan penyakit leukemia?

2. Bagaimana cara melakukan analisis *Cluster* yang telah terbentuk untuk menemukan protein signifikan dari penyakit leukemia?

### 1.3 Batasan Permasalahan

Untuk dapat memfokuskan penelitian berdasarkan latar belakang masalah dan rumusan masalah yang telah diuraikan, berikut merupakan batasan masalah yang akan diambil dalam penelitian.

1. Dataset protein penyakit leukemia yang digunakan diambil dari data *Online Mendelian Inheritance in Man (OMIM)*.
2. Penelitian membahas penyakit leukemia secara umum, tidak berfokus pada salah satu jenis leukemia akut maupun leukemia kronis.
3. Perhitungan *Centrality Measurements* yang digunakan merupakan *Betweenness, Closeness, Degree, Eigenvector*, dan *Subgraph Centrality*.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, berikut merupakan tujuan penelitian yang hendak dicapai.

1. Membentuk graf PPI berdasarkan *Centrality Measurements*.
2. Melakukan *Clustering* menggunakan algoritma *ClusterONE* pada graf PPI yang berhubungan dengan penyakit leukemia.
3. Melakukan identifikasi protein signifikan penyakit leukemia berdasarkan pada *Cluster* yang terbentuk menggunakan nilai *p-value*.

### 1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Menghasilkan *Cluster* berisikan protein signifikan terhadap penyakit leukemia.

2. Protein signifikan yang dihasilkan dapat digunakan sebagai *drug repurposing* dengan melakukan analisis senyawa lebih lanjut.
3. Hasil analisis senyawa dapat digunakan dalam pengujian laboratorium lebih lanjut untuk dijadikan obat atau terapi terhadap penderita penyakit leukemia.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN  
Berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- Bab 2 LANDASAN TEORI  
Berisi landasan teori mengenai penyakit leukemia, protein signifikan, centrality measurements, principal component analysis (PCA), dan ClusterONE.
- Bab 3 METODOLOGI PENELITIAN  
Berisi metodologi penelitian dimulai dari studi literatur, pengumpulan data, pra-proses data, pembentukan subgraf, clustering dengan ClusterONE, hingga analisis cluster.
- Bab 4 HASIL DAN DISKUSI  
Berisi penjelasan mengenai hasil penelitian yang dilakukan hingga hasil analisis cluster yang terbentuk dengan ClusterONE dan menemukan protein signifikan penyakit leukemia.
- Bab 5 SIMPULAN DAN SARAN  
Berisi tentang kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya terhadap topik yang serupa.