

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi memberikan dampak yang cukup besar pada disiplin-disiplin ilmu lainnya termasuk biologi molekuler. Sehingga menghasilkan kolaborasi ilmu yang disebut bioinformatika. Istilah bioinformatika mulai dikemukakan pada pertengahan era 1980-an untuk mengacu pada penerapan komputer dalam biologi. Namun demikian, penerapan bidang-bidang dalam bioinformatika (seperti pembuatan basis data dan pengembangan algoritma untuk analisis sekuens biologis) sudah dilakukan sejak tahun 1960-an.

Luscombe dan kawan-kawan, pada buku Essential Bioinformatika (Xiong, Jin. 2006: 4), mendefinisikan bioinformatika sebagai suatu kesatuan biologi dan informatika yang mencakup teknologi yang menggunakan komputer untuk menyimpan, mencari, memanipulasi dan mendistribusikan informasi yang berhubungan dengan mikromolekul biologi seperti genom dimana didalamnya terdiri dari DNA, RNA dan protein. Luscombe menitikberatkan pada penggunaan komputer karena sebagian besar tugas dalam analisis data genom sangat repetitif dan menggunakan matematika yang kompleks.

Genom merupakan keseluruhan bahan yang membawa semua informasi dan intruksi genetik sebuah organisme. Dengan mengekstrak dan menganalisis keseluruhan data yang terdapat pada genom, maka kita dapat mengetahui *blue print* dari suatu organisme. Contohnya pada manusia, kita dapat mengetahui warna rambut, bentuk wajah, penyakit bawaan, hingga potensi kanker yang dimiliki individu tersebut. Penyusunan genom manusia (lebih dikenal sebagai *human genome project*) dimulai pada tahun 1990 dan selesai pada tahun 2003, namun analisis lebih lanjut masih dipublikasikan hingga saat ini. Proyek penelitian tersebut tidak terbatas pada genom manusia saja, namun juga pada beberapa organisme eukariot dan prokariot. Salah satu penelitian genom pada organisme prokariot adalah pembacaan genom bakteri *Escherichia coli*.

Escherichia Coli atau lebih dikenal dengan *E. Coli* merupakan bakteri gram negatif yang banyak ditemukan pada usus hewan berdarah panas. *E. Coli* dapat membantu pembusukan makan dalam usus manusia dan juga sebagai pembentuk vitamin K untuk pembekuan darah (Riandari, Henny : 80). Kebanyakan dari *E.Coli* tidak berbahaya namun beberapa dapat menyebabkan keracunan makanan pada manusia. *E. Coli* banyak dimanfaatkan pada bidang penelitian karena dapat berkembang dengan sangat cepat dan gennya mudah untuk dimanipulasi.

Banyaknya penelitian mengenai genom seperti yang telah disebutkan diatas membuat data genom, dimana DNA, RNA dan gen termasuk didalamnya, tersedia dalam jumlah yang besar, sehingga menimbulkan suatu masalah dan tantangan baru.

Salah satu permasalahan yang timbul adalah dibutuhkan suatu alat komputasi yang mampu memprediksi lokasi dan menganalisis gen secara akurat. Banyak metode yang digunakan untuk prediksi gen. Salah satunya yang memiliki akurasi tinggi ada Hidden Markov Model.

Hidden Markov Models (HMM) adalah suatu model probabilistik yang terdiri atas *state* yang dihubungkan oleh *transitional probability*, sehingga membentuk lebih dari satu proses *Markov*. Hidden Markov Model biasanya digunakan untuk pengenalan pola. Sedangkan prediksi gen dipandang sebagai suatu pengenalan pola dari sinyal-sinyal yang membangun suatu gen. Sehingga pada penelitian ini, HMM diimplementasikan untuk memprediksi gen *Escherichia Coli*.

1.2 Rumusan Masalah

Berdasarkan uraian pada subbab diatas, dibutuhkan sebuah perangkat lunak yang mampu memprediksi dan menganalisa dengan akurats, sehingga rumusan masalah dari penelitian ini adalah bagaimana mengimplementasikan Hidden Markov Models untuk memprediksi gen pada *Escherichia Coli*.

1.3 Batasan Masalah

Pada penelitian ini diterapkan beberapa batasan masalah, yaitu :

1. Penelitian ini menggunakan *Hidden Markov Model* untuk melakukan prediksi gen *Escherichia Coli*, sedangkan untuk organisme prokariot dan eukariot lainnya tidak dibahas dipenelitian ini.
2. Prediksi gen hanya dilakukan pada prediksi *start codon*, *coding region* dan *stop codon*. Sedangkan untuk pola *ribosomal binding site* tidak dilakukan dalam penelitian ini.
3. Penelitian ini menggunakan algoritma Baum-Welch untuk *training* dan algoritma Viterbi untuk melakukan prediksi.

1.4 Tujuan Penelitian

Berdasarkan subbab 1.2 diatas maka tujuan dari penelitian ini adalah untuk mengimplementasikan *Hidden Markov Model* untuk memprediksi gen pada *Escherichia Coli*.

1.5 Sistematika Penulisan

Dalam penyusunan skripsi ini, sistematika penulisan dibagi menjadi beberapa bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan masalah secara umum meliputi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, metode penelitian dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Bab ini berisi landasan teori yang digunakan sebagai sumber dan alat dalam memahami permasalahan yang berkaitan dengan Hidden Markov Model.

BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan tentang metodologi pengembangan sistem yang digunakan dengan pendekatan terstruktur serta alat dan bahan yang digunakan dalam melakukan penelitian.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi tentang penjabaran hasil penelitian beserta pembahasan hasil penelitian tersebut.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang dapat diambil dari penelitian dan saran yang dapat dijadikan pertimbangan dalam penelitian selanjutnya.