BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kemajuan peradaban manusia berjalan seiring dengan kemajuan ilmu pengetahuan dan teknologi. Hal ini tak bisa lepas dari peran berbagai cabang ilmu pengetahuan yang mendasarinya, seperti matematika yang melahirkan berbagai cabang ilmu lain dan tentu berhubungan dengan ilmu tersebut.

Matematika tidak hanya terbatas dalam perhitungan angka saja, namun dari hal ini muncul keterkaitan yang bisa diaplikasikan dalam cabang ilmu lain. Salah satu hal yang menarik di antaranya statistika matematika yang membahas mengenai peluang yang sangat berkaitan dengan hampir seluruh peristiwa di bumi ini. Misalnya peluang kemunculan mata dadu berjumlah 1 pada pelemparan sebuah mata dadu, peluang kemunculan gambar pada saat pelemparan sebuah uang koin, peluang anak yang dilahirkan dari suatu pasangan suami istri adalah laki-laki, peluang seseorang akan sembuh setelah meminum obat tertentu jika diketahui orang tersebut sudah sakit dalam 3 hari dan lain sebagainya.

Dari contoh tersebut terdapat contoh penerapan matematika dalam kesehatan yaitu pada bidang kedokteran dan biofarmasi. Hal ini menunjukkan matematika sangat bermanfaat dan besar sekali peranannya dalam hidup manusia di antaranya dalam ilmu kedokteran dan biofarmasi.

Dalam ilmu kedokteran dan biofarmasi dikenal suatu substansi protein terpenting dari suatu makhluk hidup yang disebut dengan *Deoxyribonucleic Acid* (DNA) atau Asam Deoksiribonukleat. Substansi penting ini memiliki gugus (bagian dari substansi) protein penyusun yaitu basa nitrogen yang terdiri dari *Adenine* (A), *Guanine* (G), *Cytosine* (C) dan *Thymine* (T). Dalam kenyataannya struktur penyusun DNA ini selalu muncul dalam triplet atau barisan dari tiga macam basa nitrogen penyusun DNA. Misalnya AAA, AAG, AGG, GGG, GGC, dan lain sebagainya (terdapat $4^3 = 64$ susunan yang mungkin). Sehingga dalam penelitian seorang ahli, banyaknya basa nitrogen dari suatu DNA yang dijadikan sampel selalu kelipatan 3 (Wikipedia).

Pada perkembangan ilmu kedokteran dan biofarmasi, banyak eksperimen yang dilakukan untuk menentukan barisan basa nitrogen penyusun DNA yang digunakan dalam banyak hal. Di antaranya untuk memodelkan struktur protein pada pengembangbiakan makhluk hidup, penemuan gen (faktor penurunan sifat makhluk hidup), analisis filogenetik, pemodelan daerah DNA, subfamili (bagian dari suatu makhluk hidup yang diduga merupakan satu jenis) protein dan lainnya (Nisa Pandu, 2011).

Dalam penyelesaian tujuan-tujuan tersebut digunakan aplikasi dari perkembangan rantai Markov (Andrei A. Markov, 1906) yaitu suatu teknik peluang yang menganalisis pergerakan peluang dari suatu kondisi ke kondisi lainnya. Penelitian Hongki (2004) menggunakan rantai Markov untuk penelusuran sifat dengan menghitung peluang kromosom (kumpulan DNA) seorang anak

berasal dari leluhur ayah atau dari leluhur ibu dengan tujuan mengurangi kemungkinan penyakit bawaan pada seorang anak.

Namun penelusuran sifat hanya dilakukan secara fisik, struktur dan susunan DNA anak maupun leluhurnya tidak dipelajari dengan mendalam. Padahal diperlukan informasi apakah suatu gen berasal dari suatu mutasi atau bukan, karena jika terjadi mutasi pada gen tersebut, maka susunan DNA-nya akan berubah. Perubahan susunan DNA tersebut akan mengubah sebagian dari sifat atau karakter individu yang dikendalikan oleh gen tersebut. Sehingga para ahli melakukan pengembangan rantai Markov, yaitu Model Markov Tersembunyi atau *Hidden Markov Models* (Baum dan Petrie, 1970).

HMM saat ini sangat populer karena peranannya yang besar dalam berbagai kehidupan di antaranya dalam ilmu kedokteran dan biofarmasi. Selain itu dalam HMM banyak sekali hal yang dapat digali, dipelajari dan diteliti lebih lanjut aplikasinya dalam ilmu kedokteran dan biofarmasi. Selain untuk penelusuran sifat atau keturunan dari suatu spesies, hal yang menarik lain adalah aplikasi HMM pada DNA juga digunakan untuk menganalisis kemungkinan mutasi yang terjadi. Para ahli pada umumnya membandingkan DNA suatu spesies dengan spesies lain, misalnya perbandingan DNA manusia dengan simpanse dan lainnya. Sehingga sebelum diteliti kemungkinan terjadinya mutasi DNA, diteliti terlebih dahulu tingkat kecocokan DNA dari suatu spesies dengan DNA spesies lain.

1.2 Identifikasi Masalah

Pada sebuah model rantai markov, dikenal istilah keadaan, barisan keadaan dan perpindahan keadaan. Dalam sebuah model Markov, semua keadaan dalam suatu barisan keadaan dapat secara langsung diobservasi. Namun muncul permasalahan baru yaitu jika terdapat faktor-faktor yang tidak dapat diobservasi dan hal tersebut mempengaruhi perpindahan keadaan. Faktor-faktor tersebut dinamakan dengan keadaan yang tersembunyi. Oleh karena itu model rantai Markov yang digunakan yaitu *Hidden Markov Models* (HMM). HMM ini menggabungkan dua atau lebih rantai Markov dengan hanya satu rantai yang terdiri dari keadaan yang dapat diobservasi dan rantai lainnya yang membentuk keadaan yang tidak dapat diobservasi (tersembunyi atau *hidden*).

Ada tiga masalah utama dalam HMM, yaitu:

1. Menghitung peluang terjadinya suatu barisan observasi.
2. Menentukan keadaan tersembunyi dari suatu barisan observasi.
3. Pengestimasi parameter-parameter HMM.

Tiga masalah utama dalam HMM tersebut diselesaikan dengan beberapa algoritma diantaranya algoritma maju dan lagoritma mundur untuk menghitung peluang observasi, algoritma Viterbi untuk menentukan barisan keadaan tersembunyi, algoritma Baum-Welch dan modifikasi Baum-Welch untuk mengestimasi parameter-parameter HMM dan lain sebagainya. Pada skripsi ini akan dibahas mengenai penyelesaian masalah ketiga pada HMM, yaitu pengestimasi parameter-parameter HMM dengan menggunakan algoritma

Baum-Welch karena algoritma ini merupakan algoritma pertama yang digunakan para ahli matematika untuk mengestimasi parameter-parameter HMM.

Oleh karena itu penulis tertarik untuk menerapkan algoritma Baum-Welch dari *Hidden Markov Models* (HMM) dalam analisis kecocokan basa nitrogen penyusun DNA suatu spesies dengan DNA spesies lain. Sehingga skripsi ini diberi judul “APLIKASI ALGORITMA BAUM-WELCH UNTUK MENGESTIMASI PARAMETER-PARAMETER PADA *HIDDEN MARKOV MODELS*”.

1.3 Rumusan Masalah

Berdasarkan uraian di atas, permasalahan yang akan diangkat dalam skripsi ini dapat dirumuskan sebagai berikut:

1. Bagaimana pengestimasi parameter-parameter dari suatu barisan observasi yang diketahui pada suatu HMM menggunakan algoritma Baum-Welch?
2. Bagaimana pengaruh pengestimasi parameter-parameter HMM terhadap siklus penyelesaian masalah-masalah utama dalam HMM?
3. Bagaimana aplikasi algoritma Baum-Welch dalam mengestimasi parameter-parameter HMM dari kasus DNA makhluk hidup?

1.4 Batasan Masalah

Adapun batasan masalah dalam penulisan skripsi ini adalah sebagai berikut:

1. Perhitungan peluang observasi pada kasus DNA menggunakan algoritma maju dan mundur.

2. Penentuan barisan keadaan tersembunyi pada kasus DNA menggunakan algoritma Viterbi.
3. Analisis DNA menggunakan HMM hanya menguji tingkat kecocokan DNA suatu spesies dengan DNA spesies lain.

1.5 Tujuan Penulisan

Adapun tujuan penulisan skripsi ini adalah sebagai berikut:

1. Memahami pengestimasi parameter-parameter dari suatu barisan observasi yang diketahui pada suatu HMM menggunakan algoritma Baum-Welch.
2. Mengetahui dan memahami pengaruh pengestimasi parameter-parameter HMM terhadap siklus penyelesaian masalah-masalah utama dalam HMM.
3. Memahami aplikasi algoritma Baum-Welch dalam mengestimasi parameter-parameter HMM dari kasus DNA makhluk hidup.

1.6 Manfaat Penulisan

Analisis HMM bukanlah suatu teknik optimasi namun teknik deskriptif (gambaran) yang menghasilkan informasi peluang pada masa mendatang. Penelitian data menggunakan HMM dalam skripsi ini menambah pengetahuan akan betapa luasnya penerapan statistika matematika dalam berbagai bidang kehidupan (dalam hal ini bidang kedokteran dan biofarmasi). Selain itu penelitian

HMM ini juga dapat memberikan inspirasi untuk mengembangkan HMM dan penelitian mengenai aplikasinya dalam bidang lain.

1.7 Sistematika Penulisan

Adapun sistematika penulisan dalam skripsi ini adalah:

BAB I PENDAHULUAN

Mengemukakan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penulisan, manfaat penulisan dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Mengemukakan beberapa materi yang mendukung Bab III di antaranya menjelaskan tentang Teori Peluang, Aturan Bayes, Peubah Acak, Rantai Markov dan kajian teoritis tentang HMM dan DNA.

BAB III ALGORITMA BAUM WELCH UNTUK MENGESTIMASI PARAMETER-PARAMETER DALAM HMM

Mengemukakan kajian teoritis tentang Algoritma Baum Welch yang digunakan untuk menentukan parameter-parameter pada HMM.

BAB IV STUDI KASUS

Mengkaji aplikasi HMM dan implementasi algoritma Baum-Welch dalam kasus DNA makhluk hidup.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan dan saran hasil dari pembahasan materi.

DAFTAR PUSTAKA

LAMPIRAN