

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Berdasarkan skala global, pasar enzim dunia adalah pasar yang sangat berprospek cerah. Pemasaran enzim dunia di beberapa 3 tahun terakhir diperkirakan telah mencapai sekitar dua milyar USD (20 trilyun rupiah) dan diprediksikan pasar akan terus mengalami kenaikan rata-rata 3.3% pertahun dengan kualitas dan kualitas enzim yang meningkat (berdasar pada berita pada situs www.beritaiptek.com). Kebutuhan akan penggunaan enzim di dalam negeri bukannya tidak ada, banyak perusahaan yang bergerak dibidang industri besar yang menggunakan enzim dalam proses pengolahannya, dan kita termasuk ke dalam kategori konsumen yang sangat tergantung pada enzim impor.

Pada abad ini, salah satu penelitian yang banyak menarik perhatian dunia ilmiah sekaligus industri adalah ditemukannya mikroba (dengan enzim di dalamnya) yang mencintai lingkungan ekstrem (ekstremofil). Enzim ekstremofil yang tahan terhadap suhu ekstreme (enzim hipertermofilik), adalah enzim yang mendobrak logika ilmiah mengenai enzim sebelumnya. Hal ini didorong adanya tuntutan dunia industri yang sebagian besar prosesnya berjalan pada suhu tinggi (50-100°C). Mikroorganisme termofil menghasilkan enzim termostabil yang sangat penting dalam proses industri. Pemakaian enzim termostabil di samping tahan terhadap denaturasi panas, juga dapat meminimalkan resiko kontaminan.

Penggunaan enzim hipertermofilik banyak digunakan pada industri pembuat pemanis, industri kertas dan pulp, dan untuk proses bukti uji DNA (*finger printing*

DNA). Pemakaian enzim impor oleh banyak perusahaan jika ditilik dari segi ekonomi, jauh lebih murah dibandingkan dengan membeli produk enzim domestik yang hanya bisa diproduksi secara skala kecil dan kualitas yang kurang diminati. Kendala untuk memproduksi enzim di dalam negeri umumnya disebabkan oleh dibutuhkannya dana yang besar karena bahan-bahan pereaksinya hampir semuanya didatangkan dari luar negeri serta experiment yang terkait dengan pengerjaan secara manual di laboratorium untuk meningkatkan kualitas enzim, akan melibatkan kurun waktu yang lama (proses *trial and error*).

Salah satu experiment laboratorium yang memakan waktu lama adalah penelitian mengenai mutasi gen pada enzim. Mutasi gen ini dilakukan untuk mendapatkan hasil akhir dari enzim yang memenuhi harapan (termofilik, stabil). Tingkat kesukaran untuk memutasi suatu asam basa amino dalam *sequence* enzim sangat tinggi dikarenakan faktor seperti pertimbangan terhadap pemilihan dari asam amino yang akan dimutasi (jenisnya dan banyaknya), prosedural yang belum baku (*trial and error*), dan kondisi pengerjaan yang rentan kali gagal. Selama ini percobaan untuk berusaha mendapatkan kualitas enzim yang sesuai melalui mutasi gen hanya berdasar kepada perbandingan terhadap sejumlah literatur tertentu, dan umumnya pada saat pengerjaan di laboratorium, hasil yang di dapat kadangkala tidak sesuai. Berdasarkan pada pengalaman penulis dalam mengerjakan penelitian di laboratorium, dana yang dibutuhkan untuk melakukan experiment tergolong besar dan pengerjaan manual di laboratorium yang lama (sekitar 2-3tahun).

Seiring dengan perkembangan teknologi yang bergerak di bidang informatika, pengerjaan experiment secara manual dapat diminimalisir kesalahannya dengan mengekstrak informasi dari data pada GenBank yang dapat

diakses secara bebas melalui koneksi internet dan tanpa adanya pembayaran untuk lisensi. Hasil informasi yang didapat akan dipergunakan untuk penarikan prediksi yang akan membawa pada interpretasi kemungkinan dari keberhasilan experiment di laboratorium. Peleburan antara ilmu informatika untuk mengolah data biologikal dinamakan dengan Bioinformatika.

Untuk mendapatkan enzim yang bersifat thermostabil beberapa penelitian dilakukan. Penelitian yang secara langsung diketahui oleh penulis antara lain adalah penelitian dengan tajuk *Improvement of the amylase stability by protein engineering* yang dilakukan oleh Prof. Dr. Soetijoso Soemitro dan Dr. Dessy Natalia dalam proyek Hibah Bersaing IX, 2000-2003, *Production of thermostable α -amylase by using Pichia pastoris as the host* oleh Prof. Dr. Soetijoso Soemitro dan Dr. Dessy Natalia dalam proyek Riset Unggulan Terpadu XII, 2004-2006.

Salah satu hal yang diduga dapat mempengaruhi termostabilitas enzim adalah melalui mutasi asam amino yang menyebabkan perubahan dari perpanjangan urutan basa nukleotida hasil mutasi. Dengan memanfaatkan database dari Genbank yang dapat diakses bebas dari situs <http://www.ncbi.nlm.nih.gov/>, penggalian informasi yang didapat dari hasil me-mining diperkirakan akan dapat memprediksi mengenai pola perubahan sifat termostabilitas enzim yang termutasi. Berangkat dari beberapa acuan *literature* mengenai data *mining* dalam bidang bioinformatika, melalui penelitian ini penulis mencoba untuk dapat me-mining informasi dari tumpukan data yang ada pada Genbank dengan menggunakan metode BLAST (*Basic Local Alignment Tools*) dan SVM (*support vector machine*).

1.2 PERUMUSAN MASALAH

Selama ini experiment di laboratorium selalu dilakukan secara manual dan selalu menggunakan prosedural yang masih bersifat *trial and error* (coba-coba), sehingga tidak adanya prosedural yang akurat untuk menghasilkan produk akhir yang diharapkan. Waktu pengerjaan yang lama dan dana yang terpakai juga cukup besar tetapi kegagalan perolehan hasil akhir selalu ada. Tidak adanya suatu pola prediksi yang dapat dijadikan sebagai acuan untuk mengerjakan experiment menjadi salah satu penyebab kegagalan tersebut.

Salah satu penelitian yang rentan terhadap kegagalan adalah, mengenai mutasi gen. Untuk dapat menghasilkan suatu produk enzim baru dengan kualitas baik salah satu caranya dengan memutasikan urutan asam amino pada enzim tersebut, sehingga produk akhir berupa enzim yang termostabil dapat diperoleh.

SVM (*support vector machine*) adalah suatu *machine learning* yang telah diaplikasikan pada beberapa penelitian bioinformatika, penelitian yang menggunakan SVM antara lain : untuk memprediksi sisi glikosilasi protein (Caragea,C *et al.*,2007); pengenalan *Translation Initiation Sites* (TIS) pada protein/enzim (Zien,A *et al.*, 2000); penerapan SVM untuk *drug design* (Burbidge,R *et al.*, 2001). Penerapan SVM untuk memprediksi kestabilan protein karena pengaruh single mutation (Gromiha MM, *et al.* 2000 dan Guerois R *et al.* 2002).

Penerapan metode SVM ini digunakan untuk membantu dalam hal memprediksi batasan-batasan tertentu yang dapat diolah lebih lanjut dalam pengambilan keputusan untuk menentukan arah experiment yang akan dilakukan. Penelitian yang dilakukan oleh penulis merupakan tindak lanjut dari penelitian yang

dilakukan oleh Gromiha MM, *et al* (2000) dan Guerois R *et al.* (2002), dimana pada penelitian sebelumnya, mereka menggunakan metode SVM untuk menjawab permasalahan mengenai pengaruh kestabilan protein karena single mutation yang terjadi. Penulis ingin menyelidiki lebih lanjut mengenai “kemampuan” metode SVM untuk menganalisa kestabilan protein karena adanya multiple mutation.

Berdasarkan hal tersebut, Beberapa masalah yang dapat diidentifikasi pada penelitian ini antara lain:

1. Korelasi antara perpanjangan asam amino yang di mutasi terhadap perubahan termostabilitas enzim mutan.
2. Apakah metode SVM dapat memberikan keakuratan solusi untuk menjawab pengaruh dari *multiple mutation* terhadap termostabilitas protein.
3. Apakah penerapan TI dapat membantu dalam pengambilan keputusan dalam penentuan experiment laboratorium yang umumnya selalu dilakukan secara manual.

1.3 TUJUAN DAN MANFAAT PENELITIAN

Tujuan penelitian ini adalah untuk menganalisa sejauh mana pengaruh perpanjangan mutasi suatu protein terhadap perubahan termostabilitasnya dengan menggunakan metode SVM.

Manfaat penelitian ini adalah :

- Memberikan solusi terhadap experiment di laboratorium yang dapat menghabiskan waktu dan biaya yang mahal.

- Menerapkan peranan IT untuk membantu permasalahan dalam laboratorium.
- Menemukan korelasi antara jenis asam amino dan perpanjangan mutasi asam amino terhadap kestabilan dan perubahan sifat biologi dari suatu enzim.
- Sebagai acuan untuk melakukan penelitian lebih dalam mengenai enzim dengan metode *pattern recognition* oleh SVM.

1.4 RUANG LINGKUP

Dalam penulisan proposal tesis ini, masalah yang akan dianalisis dibatasi sebagai berikut:

1. Persiapan *Dataset*, yaitu tahap pengambilan data dan seleksi data.
 - Proses pengambilan data *sequence* enzim dari GenBank yang berasal dari situs NCBI (*National Center for Biotechnology Information*) yang dapat diakses dan digunakan secara bebas tanpa adanya biaya untuk membayar lisensi (*Use of [National Center for Biotechnology Information, U.S. National Library of Medicine](http://www.ncbi.nlm.nih.gov/) Central requires no registration or fee. Access it from any computer with an Internet connection*). Pengguna dapat men-*download* kebutuhan informasi mengenai biologi molekular dari NCBI di situs <http://www.ncbi.nlm.nih.gov/>.
 - Seleksi data dengan menggunakan metode BLAST *all-against-all*, parameter data yang diambil berdasar pada presentasi homolog >70%.

2. Transformation, yaitu proses transformasi data kedalam *high dimensional-space* dengan fungsi kernel yang menghasilkan data *mapping* dalam ruang vector.
3. Training *Dataset*, yaitu proses pengolahan data training set untuk menemukan garis pemisah (*separating hyperplane*) kelas berdasarkan tolak ukur "margin".
4. testing *Dataset*
5. SVM, yaitu proses untuk menganalisa data dan mencari *hyperplane* yang terbaik untuk mengelompokan data menjadi dua class data untuk mendapatkan *pattern*.
6. Interpretation, yaitu proses untuk membaca hasil *scoring function* data untuk dilakukan prediksi hasil.

1.5 SISTEMATIKA PENULISAN

Sistematika penulisan ini ditujukan untuk memberikan gambaran mengenai hal-hal yang akan dibahas dalam tesis ini. Adapun sistematika penulisan disajikan sebagai berikut :

Bab I : Pendahuluan

Bab ini berisikan mengenai latar belakang, tujuan penelitian, ruang lingkup/ pembatasan masalah, metode penelitian, dan sistematika penulisan.

Bab II : Teori Data Mining dan SVM (*Support Vector Machine*)

Bab ini membahas mengenai teori-teori penunjang yang berhubungan dengan permasalahan yang dihadapi dalam topik tesis ini. Teori-teori yang dibahas meliputi teori Data mining dan SVM (*Support Vector Machine*)

Bab III : Penerapan TI Pada *Laboratorium Research*

Bab ini membahas mengenai teori-teori penunjang yang berhubungan dengan pemanfaatan TI untuk penelitian di laboratorium yang berhubungan dengan DNA rekombinan (mutasi gen) serta membahas mengenai teori Enzim dan Bioinformatika.

Bab IV: Metode Penelitian

Pada bab ini dijelaskan mengenai langkah-langkah penelitian yang dilakukan dalam menyelesaikan permasalahan dan metode-metode yang digunakan dalam setiap tahapan penelitian.

Bab V : Hasil Pengujian Sistem dan Pembahasan

Pada bab ini dijelaskan mengenai hasil pengujian yang didapat. Pada penelitian ini hasil pengujian dikategorikan berdasarkan modul yang ada pada sistem.

Bab VI: Kesimpulan dan Saran.

Bab ini berisi kesimpulan dan saran penulis mengenai penelitian ini.