

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 dikenal dengan 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 dalam pengembangan algoritma untuk analisis sekuens biologis) sudah dilakukan sejak tahun 1960-an. Bioinformatika ialah ilmu yang mempelajari penerapan teknik komputasi untuk mengelola dan menganalisis informasi hayati. Bidang ini mencakup penerapan metode-metode matematika, statistika, dan informatika untuk memecahkan masalah-masalah biologi, terutama yang terkait dengan penggunaan sekuens DNA dan asam amino. Contoh topik utama dalam bidang ini meliputi pangkalan data untuk mengelola informasi hayati, penyejajaran sekuens (*sequence alignment*), prediksi struktur untuk meramalkan struktur protein atau pun struktur sekunder RNA, analisis filogenetik, dan analisis ekspresi gen (Rahayu & Nugroho, 2015). Salah satu penerapan ilmu bioinformatika adalah pada bidang pertanian yaitu dengan memanfaatkan gen pada tumbuhan untuk mendapatkan variatas unggul.

MAIZE (*Zea mays L.*) atau jagung adalah tanaman penting di seluruh dunia dalam hal produksi biji-bijian dan *stover* dengan sekitar 36 juta ha ditanam pada tahun 2010 di Amerika Serikat (USDA & NASS, 2011). *Zea mays* (jagung) memiliki produksi biji gandum tertinggi di dunia, menghasilkan 875 juta ton pada tahun 2012 (www.faostat.fao.org). Jagung juga merupakan komoditas strategis bagi Indonesia karena mempunyai dimensi penggunaan yang luas seperti pakan ternak (langsung atau olahan), pangan pokok bagi sebagian penduduk (berpotensi untuk masyarakat yang lebih luas) dan jajanan, bahan baku industri (pati, gula, pangan olahan), dan energi (bioetanol).

Sepuluh dari penggunaan saat ini adalah sebagai bahan baku utama industri pakan ternak. Penggunaan lain meliputi bahan pangan langsung, bahan baku minyak nabati non kolesterol, tepung jagung dan makanan kecil. Pengembangan jagung harus melihat potensi dan struktur kebutuhan tersebut secara komprehensif (Ditjentan, 2010).

Pada beberapa tahun terakhir, kebutuhan jagung terus meningkat seiring dengan semakin meningkatnya laju pertumbuhan penduduk dan peningkatan kebutuhan untuk pakan, apalagi dengan bertambahnya produksi bioetanol sebagai bahan bakar. Hal ini karena jagung merupakan sumber karbohidrat, protein, serat, dan lemak. Komposisi inilah yang menjadikan jagung sebagai bahan pangan pokok yang prospektif karena kandungan karbohidratnya mendekati beras (78,9%). Oleh karena itu, jagung adalah komoditas yang mempunyai kegunaan yang luas baik pangan, pakan dan energi (Bantacut, 2010).

Selain nilai ekonominya, jagung adalah organisme model penting untuk studi di genetika tanaman, fisiologi, dan pengembangan. Jagung dibedakan dari rumput lain karena genomnya berasal dari peristiwa tetraploidi kuno yang unik bagi garis keturunannya. Banyak faktor yang mempengaruhi produksi jagung yang meliputi praktek budidaya, pemuliaan tanaman (pembibitan oleh petani, hibrida, dan perbaikan jarak tanam) (Duvick, 2005), pola tanam (Cook, 2006), pemupukan (Ma dan Dwyer, 1998), pengendalian hama (Christou *et al.*, 2006), pengairan (Payero *et al.*, 2006), serta panen dan pasca panen (Magan and Aldred, 2007), penyimpanan dan distribusi (Shapiro *et al.*, 1989), dan pengolahan (Rausch and Belyea, 2006). Selain produksi jagung yang meningkat kendala yang di alami petani adalah produksi akan berkurang bila pada tanaman jagung terdapat gulma, serangan hama, penyakit, tanaman pengolahan dan pemeliharaan yang kurang optimal, penanaman varietas yang berproduksi rendah dan kekeringan (Suprpto, 1995).

Upaya perbaikan sifat tanaman dengan meningkatkan keragaman genetiknya perlu dilakukan. Seperti telah diketahui, modal dasar pemuliaan tanaman adalah adanya keragaman yang luas. Dengan adanya variabilitas yang luas, proses seleksi dapat dilakukan secara efektif karena akan memberikan

peluang yang lebih besar untuk diperoleh karakter-karakter yang diinginkan (Sobir, 2007). Dapat dikatakan bahwa syarat utama yang diperlukan oleh pemulia untuk merakit varietas unggul baru adalah tersedianya materi genetik dengan keragaman yang luas. Keragaman genetik di alam timbul dari gen-gen yang bersegregasi dan berinteraksi dengan gen lain melalui hibridisasi, mutasi, dan introduksi (Crowder, 1988).

Mengingat kompleksitas dan volume data biologis besar, dibutuhkan sebuah pendekatan komputasi modern yang disebut *machine learning*. *Machine learning* telah menjadi alat yang sangat penting untuk proyek dalam biologi komputasi, bioinformatika, dan informatika kesehatan. *Machine learning* memiliki kinerja seperti otak manusia, dimana dibutuhkan sebuah pembelajaran agar dapat bekerja sesuai dengan kehendak. Algoritma *machine learning* telah digunakan untuk pelatihan dan pengenalan fitur utama dan untuk klasifikasi kelompok. Kekuatan metode *machine learning* adalah dapat mendeteksi pola yang sulit dilihat dari kumpulan data yang besar, *noisy* atau kompleks. Berdasarkan hal tersebut, cara kerja seperti itulah yang dibutuhkan sebagai alat yang mendukung dalam penelitian ini. Salah satu metode *machine learning* yang berkembang saat ini serta mendukung proses pengolahan dalam bidang bioinformatika adalah *Support Vector Machine* (SVM). SVM sangat kuat dalam mengenali pola halus dalam kumpulan data kompleks (Priya, 2013).

SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*. *Support Vector Machine* (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. Sebagai salah satu metode *pattern recognition*, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai *state of the art* dalam *pattern recognition*, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat (Nugroho, 2003).

1.2 Rumusan Masalah

Berdasarkan latar belakang, permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut :

1. Bagaimana aplikasi metode SVM pada data *maize growth stage* ?
2. Bagaimana pengklasifikasian terhadap data uji dengan menggunakan metode SVM ?
3. Bagaimana melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan parameter akurasi SVM ?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah :

1. Data yang digunakan adalah data *maize growth stage* dari http://qteller.com/rna_data_sources.php dan <https://www.ncbi.nlm.nih.gov/bioproject?term=PRJNA343268&cmd=DetailsSearch> diakses pada tanggal 28 Mei 2018.
2. Bahasa pemrograman dan *software* yang digunakan dalam penelitian ini adalah *software RStudio* dengan beberapa *package bioinformatics* dari *bioconductor*.
3. Melakukan klasifikasi dengan menggunakan metode SVM linear.

1.4 Jenis Penelitian dan Metode Penelitian

Jenis penelitian yang dilakukan peneliti adalah kategori aplikasi dengan menggunakan metode SVM pada data bioinformatika

1.5 Tujuan

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, tujuan penelitian ini adalah :

1. Mengetahui pengaplikasian metode SVM pada data *maize growth stage*
2. Melakukan pengklasifikasian terhadap data uji dengan menggunakan metode SVM
3. Melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan parameter akurasi SVM

1.6 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah :

1. Hasil penelitian ini diharapkan dapat digunakan oleh berbagai kalangan terutama untuk disiplin ilmu bioinformatika yang didalamnya termasuk ilmu-ilmu hayati (kimia, biologi, farmasi, dan kedokteran) seperti mengidentifikasi yang mempengaruhi kesalahan dalam pengklasifikasian dan mengetahui fungsi dari nama gen yang memiliki bobot tertinggi, serta untuk ilmu fisika, statistika dan teknologi informasi seperti penggunaan metode SVM.
2. Hasil penelitian ini diharapkan mampu memberikan informasi kepada berbagai kalangan khususnya disiplin ilmu biologi dan statistika terkait pengelompokkan dan dapat memprediksi kelompok dari data yang baru.