

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi saat ini menyumbang pertumbuhan data yang sangat pesat. Arus data dan informasi tumbuh signifikan dalam ukuran jumlah dan media yang bervariasi, dengan sebutan era *Big Data*. Big data adalah sebuah sistem teknologi yang diperkenalkan untuk menanggulangi ‘ledakan informasi’ seiring dengan semakin bertumbuhnya ekosistem pengguna perangkat *mobile* dan data *internet*. Big data memiliki tiga istilah yaitu *volume*, *variety*, dan *velocity*. *Volume* berkaitan dengan ukuran media penyimpanan data yang sangat besar atau mungkin tak terbatas, *variety* berarti tipe atau jenis data yang dapat diakomodasi, dan *velocity* dapat diartikan sebagai kecepatan proses (Bayu, 2018). Teknologi *Big Data* dapat dimanfaatkan di berbagai bidang termasuk biologi molekular yang melahirkan bidang bioinformatika. Istilah bioinformatika berkembang pesat pada pertengahan tahun 1990, yang disebabkan berkembangnya teknologi sekuen DNA. Bioinformatika adalah aplikasi ilmu yang mempelajari bidang komputer untuk mengelola dan menganalisis informasi biologis. Bioinformatika mencakup aplikasi metode-metode statistika, dan informatika untuk memecahkan persoalan biologis (Attwood, 1999).

Bioinformatika sejak lama telah digunakan untuk mendukung eksperimen pada bidang pertanian. Salah satu tanaman penting dan strategis dalam upaya pembangunan pertanian di Indonesia adalah jagung. Tanaman yang memiliki nama latin (*Zea mays L.*) ini merupakan salah satu tanaman pokok bagi kebutuhan manusia. Peran jagung di Indonesia digunakan sebagai bahan pangan telah beralih Hampir 50% kebutuhan jagung nasional digunakan untuk industri ternak (Widiyanti, 2016). Jagung tidak hanya digunakan sebagai bahan pangan dan pakan saja, tetapi juga digunakan sebagai bahan baku industri lainnya, seperti bahan bakar alternatif (biofuel), *polymer* dan lain-lain. Permintaan jagung baik untuk industri pangan, pakan, dan kebutuhan industri lainnya dalam lima tahun ke

depan di proyeksikan akan terus meningkat seiring dengan terus bertambahnya jumlah penduduk dan juga peningkatan pendapatan dan daya beli masyarakat (Kementan, 2017).

Kebutuhan jagung yang semakin banyak, mendorong berbagai usaha dengan tujuan untuk meningkatkan, mempertahankan dan memperoleh sifat-sifat tumbuhan yang baik sehingga memperoleh hasil maksimal. upaya ini sering disebut sebagai pemuliaan tanaman (Allard, 1960). Seleksi yang akurat terhadap suatu karakter yang diinginkan dari tanaman adalah dengan berdasarkan pada gen yang mengendalikan karakter tersebut. Untuk kepentingan ini maka informasi genetik dari suatu tanaman khususnya yang terkait dengan suatu karakter sangatlah penting. Untuk itu maka identifikasi genetik dengan pendekatan molekuler sangat dibutuhkan dalam kegiatan pemuliaan ini agar memperoleh hasil yang tepat (Nuraida, 2012).

Pendekatan molekuler telah menghasilkan suatu alat bantu untuk keperluan pemuliaan tanaman. Parikesit, dkk (2017) mengatakan baru-baru ini, hadirnya teknologi *Next Generation Sequencing* (NGS) telah mengubah paradigma dalam penelitian berbasis bioinformatika. Penelitian berbasis bioinformatika telah digunakan dalam (1) pengembangan penanda molekuler (Priyono & Putranto, 2014); (2) desain primer untuk analisis ekspresi gen diferensial (Budiani et al. 2016); (3) pengembangan peta genetik (Priyono & Putranto, 2016); dan (4) analisis ekspresi gen (Putranto et al. 2015). Teknologi NGS memungkinkan pengumpulan data ekspresi gen skala besar dengan biaya terjangkau dalam waktu singkat.

Mengingat kompleksitas dan besarnya volume data biologis, sehingga diciptakan sebuah pendekatan komputasi modern yang disebut *machine learning*. *Machine learning* telah menjadi alat yang penting untuk proyek dalam biologi komputasi, bioinformatika, dan informatika kesehatan. *Machine learning* bekerja 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, *Machine learning* berguna sebagai alat pendukung dalam penelitian ini. Salah satu metode *machine learning* yang berkembang saat ini serta mendukung proses pengolahan dalam bidang bioinformatika adalah *biclustering*.

Pada penelitian ini data yang digunakan adalah ekspresi gen, pada *clustering* pengelompokan setiap gen didefinisikan menggunakan semua kondisi, dan pengelompokan kondisi dicirikan oleh aktivitas semua gen. Sehingga yang didapat adalah *cluster* global, hal ini menyebabkan banyak pola aktivasi gen hanya dibawah kondisi eksperimental tertentu dan banyak informasi yang dikesampingkan. Mengatasi keterbatasan metode *clustering*, pada analisis *biclustering* dilakukan pengelompokan secara simultan pada dua dimensi matriks data, yaitu pada baris dan kolom. Dengan menggunakan metode ini, akan terungkap submatriks dengan elemen-elemen yang memanifestasikan perilaku yang sama dari anggota kelompok. Submatriks ini terdiri dari subset kolom yang bisa mencirikan suatu kelompok baris. Memahami gen dan sampel dari tanaman jagung, diharapkan berguna dalam proses pengembangan pemuliaan tanaman jagung.

## 1.2 Rumusan Masalah

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

1. Bagaimana aplikasi algoritma *CC biclustering* pada data *Maize Growth Stages*?
2. Bagaimana identifikasi pola data dengan menggunakan algoritma *CC biclustering*?
3. Bagaimana implikasi hasil *bicluster* dengan menggunakan algoritma *CC biclustering* pada data *Maize Growth Stages*?

## 1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Melakukan analisis *CC biclustering* pada data ekspresi gen.

2. Data yang digunakan adalah data *Maize Growth Stages* dari alamat website [http://qteller.com/rna\\_data\\_sources.php](http://qteller.com/rna_data_sources.php) dan website NCBI yaitu <http://www.ncbi.nlm.nih.gov/bioproject?term=PRJNA343268&cmd=DetailsSearch> yang diakses pada bulan April 2018.
3. Bahasa pemrograman dan *software* yang digunakan dalam penelitian ini adalah *software* RStudio dengan beberapa *package bioinformatics* dari *bioconductor*.

#### **1.4 Tujuan**

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

1. Mengetahui pengaplikasian algoritma *CC biclustering* pada data *Maize Growth Stages*.
2. Mengidentifikasi pola data dengan menggunakan algoritma *CC biclustering*.
3. Mengetahui implikasi hasil *bicluster* dengan menggunakan algoritma *CC biclustering* pada data *Maize Growth Stages*.

#### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah:

1. Secara teoritis  
Hasil penelitian ini diharapkan dapat digunakan oleh berbagai kalangan terutama untuk disiplin ilmu bioinformatika dalam menambah wawasan dan pengetahuan.
2. Secara praktek  
Hasil penelitian ini diharapkan mampu memberikan informasi kepada berbagai kalangan terkait pengetahuan gen dan organ pada jagung yang koheren dan dapat digunakan untuk pengetahuan dalam pemuliaan tanaman.