

BAB I

PENDAHULUAN

1.1. Latar Belakang

Etanol (*alcohol*) adalah nama suatu golongan senyawa organik yang mengandung unsur C,H dan O. Etanol dalam ilmu kimia disebut juga sebagai etil alkohol C_2H_5OH . Rumus umum dari *alcohol* adalah $R - OH$. Secara struktur *alcohol* sama seperti air, namun salah satu gugus hidrogennya digantikan oleh gugus alkil. Gugus fungsional *alcohol* adalah gugus hidroksil, OH. Pemberian nama *alcohol* biasanya dengan menyebut nama alkil yang terikat pada gugus OH, kemudian menambahkan nama alkohol (Siregar, 1998)

Menurut data yang diperoleh dari *World Health Organization* (WHO) memperkirakan saat ini jumlah pecandu alkohol diseluruh dunia mencapai 64 juta orang (Maula and Yuniastuti, 2017). Di Indonesia, menurut lembaga Badan Pusat Statistika total konsumsi alkohol per kapita (15+) dalam liter perkapita pada 2017 sebesar 0.27 untuk wilayah perkotaan dan sebesar 0.85 untuk wilayah perdesaan. Badan Narkotika Nasional (BNN) memperkirakan ada 3,2 juta orang (1,5% dari total populasi) di Indonesia mempunyai riwayat menggunakan NAPZA diantaranya 46% adalah perilaku minum alkohol (Triyono, 2014). Data dinas penelitian dan pengembangan Dislitbang Polri (2014), pengguna alkohol remaja mulai dari usia 14-16 tahun (47,7%), 17-20 tahun (51,1%) dan 21-24 tahun (31%).

Konsumsi etanol berdampak pada metabolisme manusia, menurunkan glukosa darah dengan menstimulasi respon sekresi insulin awal yang distimulasi glukosa, yang dapat menyebabkan hipoglikemia sementara (McMonagle and Felig, 1975). Konsumsi etanol dapat mempengaruhi sampai pada tingkat gen dalam tubuh manusia, umumnya pengaruh yang didapat berbeda antara individu dengan individu yang lain. Pada beberapa kasus konsumsi etanol menyebabkan kehilangan kesadaran dan bahkan

menyebabkan kecelakaan yang berakibat kematian (Marsaid and Hidayat, 2013) untuk beberapa kasus yang lain masih ditemukan nilai konsentrasi *alcohol* dalam gen darah setelah beberapa hari konsumsi, sehingga efek pengonsumsi *alcohol* tidak bisa hilang begitu saja setelah beberapa hari konsumsi (Kupfer, 2013). Sehingga dalam penanganan pasien yang terpapar oleh *alcohol* perlu dilakukan penelitian untuk mengetahui pada gen apa yang harus dilakukan penanganan secara medis (Kupfer, 2013). Maka dari itu perlu untuk dilakukan adanya pengelompokan gen agar penanganan secara medis tepat sasaran pada pasien, untuk mengetahui kelompok karakteristik pengaruh dalam tingkat ekspresi gen, dilakukan dengan menggunakan RNA yang diekstraksi dari seluruh darah, melintasi jalur waktu lima poin ketika etanol memasuki sistem darah, mencapai tingkat 0,08 g / dL (0,08% BAC), dan kembali ke 0,02% BAC, konsentrasi terendah dari deteksi Breathalyzer (Kupfer, 2013).

Dalam melakukan analisis karakteristik pada tingkatan gen di perlukan disiplin ilmu khusus yaitu bioinformatika. Bioinformatika adalah gabungan antara ilmu biologi dan ilmu teknik informasi (TI). Pada umumnya bioinformatika didefinisikan sebagai aplikasi dari alat komputasi dan analisa untuk menangkap dan menginterpretasikan data-data biologi. Ilmu ini merangkup berbagai disiplin ilmu termasuk ilmu komputer, matematika dan fisika, biologi, dan ilmu kedokteran. Istilah bioinformatik mulai diperkenalkan pertengahan era 1980-an untuk mengacu pada penerapan komputer dalam biologi (Fatchiyah, 2009). Salah satu teknologi yang berkembang dalam bioinformatik adalah teknologi *microarray*. *Microarray* merupakan teknologi dalam bidang Biologi Molekuler dan Medis yang dapat digunakan untuk melihat perbedaan ekspresi gen. Selain itu, *microarray* dapat digunakan untuk mendeteksi *single nucleotide polymorphism* (SNP) and *genotyping*. Teknologi ini memanfaatkan kumpulan *array* yang berjumlah ribuan yang berisi nukleotida DNA yang berfungsi sebagai *probe*. Eksperimen *microarray* dapat dikatakan sebagai tes genetik yang dilakukan secara paralel (Biogen Litbang Pertanian, 2010). Dengan analisis *microarray*,

perubahan ekspresi gen dan evaluasi gen yang diminati dapat identifikasi dengan set gen yang berfungsi sebagai penanda biologis potensial untuk efek terkait alkohol.

Salah satu metode analisis yang dipelajari dalam ilmu *bioinformatic* adalah *machine learning*. *Machine learning* adalah satu disiplin ilmu yang mempelajari bagaimana membuat komputer yang memiliki kecerdasan buatan. *Machine learning* berfungsi untuk mengoptimalkan performa dari sistem dengan menggunakan data sampel atau data histori. *Machine learning* memiliki model yang ditetapkan dengan beberapa parameter, untuk mengoptimalkan parameter model tersebut dilakukan dengan menggunakan data pelatihan atau data histori. Beberapa metode analisis dalam *machine learning* adalah *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Alpaydin, 2010). Informasi mengenai hasil pengelompokan karakteristik gen yang terpapar oleh *alcohol* dapat dilakukan dengan menggunakan ilmu *bioinformatic* dengan salah satu metode analisis *machine learning* yaitu *unsupervised learning*. Dalam metode *unsupervised learning* salah satunya dikenal metode *biclustering* (Alpaydin, 2010).

Biclustering adalah sebuah metode yang bertujuan untuk mempartisi elemen menjadi set (*cluster*), dengan mengoptimalkan homogenitas grup (yaitu, elemen *cluster* harus sangat mirip satu sama lain), dan pemisahan grup, elemen dari berbagai *cluster* harus memiliki kesamaan yang rendah satu sama lain secara Bersama sama antara baris dan kolom. *Biclustering* dapat digunakan untuk mendeteksi subkelompok yang relevan terhadap struktur tersembunyi dari sebuah data. *Biclustering* dapat mengkarakterisasi fenomena biologis yang diwujudkan oleh kumpulan *biclusters*, masing-masing *bicluster* mewakili jenis perilaku gabungan yang berbeda dari satu set gen dalam set sampel yang sesuai (Piscopo and Marina, 2017).

Pada metode *biclustering* spektral menggunakan pendekatan aljabar linier, dengan menggunakan *Singular Value Decomposition* (SVD). Menggunakan SVD, matriks data D dimensi $N \times M$ dapat didekomposisi sebagai $D = U \Lambda V^T$, Metode *biclustering* spektral menggabungkan matriks

normalisasi yang terintegrasi (Kluger, 2003). Pendekatan tersebut dilakukan dengan membuang data yang tidak relevan latar belakang penelitian, sehingga tidak ada argumen tambahan diperlukan untuk mengabaikan kontribusi yang terkait dengan nilai eigen terbesar, seperti yang dianjurkan dalam (Alter, Brown and Botstein, 2000). Sehingga analisis *spectral biclustering* di anggap sesuai untuk mengetahui pengelompokan nilai ekspresi gen yang terpapar oleh etanol

Maka dari itu, sesuai dengan hal-hal yang telah disampaikan sebelumnya, penelitian kali ini ingin mendeteksi *bicluster* pada data *characterization of gene expression changes in blood during acute* sebagai bahan kajian untuk mendeteksi kelompok karakteristik dalam ekspresi gen menggunakan analisis model *spectral* dengan alat bantu bahasa pemrograman R.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka penulis dapat merumuskan beberapa masalah yang hendak diselesaikan pada penelitian ini, antara lain:

1. Bagaimana hasil eksplorasi data *Characterization of gene expression changes in blood during acute* menggunakan bahasa pemrograman R?
2. Bagaimana hasil penelompokan data *Characterization of gene expression changes in blood during acute* menggunakan *bicustering* dengan algoritma *Spectral* menggunakan bahasa pemrograman R?
3. Bagaimana mengimplementasikan hasil kajian algoritma *Spectral* dengan bahasa pemrograman R tersebut pada program kesehatan di Indonesia?

1.3. Batasan Masalah

Data yang digunakan pada penelitian ini adalah data *characterization of gene expression changes in blood during acute* (GSE20846). Pada penelitian ini menggunakan bahasa pemrograman R versi 3.5.2 dengan *package.Biclust* untuk analisis *biclustering*.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui informasi hasil eksplorasi data *Characterization of gene expression changes in blood during acute* dengan bahasa pemrograman R..
2. Mengetahui hasil pengelompokan data *Characterization of gene expression changes in blood during acute* menggunakan *bicustering* dengan metode *Spectral* menggunakan bahasa pemrograman R.
3. Mengetahui implementasi hasil kajian *bicustering* dengan bahasa pemrograman R tersebut pada program kesehatan di Indonesia.

1.5. Manfaat Penelitian

1. Mengetahui hasil pengaplikasian *bicustering* algoritma *spectral* pada data *characterization of gene expression changes in blood during acute*.
2. Dapat mengetahui hasil pengelompokan terhadap data uji dengan menggunakan metode *bicustering* algoritma *spectral*.