

1. Pendahuluan

Pada bagian pendahuluan memuat beberapa sub bab. Sub bab tersebut diantaranya latar belakang, perumusan dan batasan masalah, tujuan, dan organisasi penulisan. Latar belakang mengemukakan masalah dan rencana penelitian, perumusan dan batasan masalah mengemukakan permasalahan dan batasan pada penelitian, sub bab tujuan mengemukakan tujuan yang dicapai pada penelitian, dan organisasi penulisan mengemukakan struktur penulisan pada *paper* ini.

Latar Belakang

Bioinformatics merupakan bidang area yang mengaplikasikan molekular biologi ke dalam data yang berisi rangkaian DNA manusia (*genomic data*) dengan teknologi komputer. Teknologi ini membantu peneliti untuk mendiagnosa masalah kesehatan secara akurat dengan mengukur level mRNA akan didapatkan informasi yang terjadi dengan keadaan sel sehingga dapat memahami informasi proses biologikal yang terjadi didalamnya, contohnya adalah penyakit kanker. Kanker merupakan salah satu penyakit yang mematikan di dunia [1]. Salah satu cara untuk menghindari penyakit ini adalah dengan mendeteksi gejalanya sejak dini. Data kanker dapat dipresentasikan dengan menggunakan teknologi DNA *microarray*.

DNA *microarray* merupakan sebuah *chip* yang terbuat dari silikon/kaca seperti dalam *Affymetrix array* atau *microscopic head* dalam *Illumina array* dimana ribuan untaian molekul atau *oligonucleotides complimentary DNA (cDNA)* ditanamkan yang dinamai dengan *feature*. Dengan membandingkan hasil dari teknologi DNA *microarray* dengan sel yang normal, peneliti dapat mengukur perubahan yang terjadi pada level ekspresi gen sehingga semua ribuan ekspresi gen dapat diinvestigasi secara bersamaan [2].

Untuk mendeteksi gejala kanker, salah satu teknik yang dapat digunakan adalah melakukan *data mining* terhadap *feature* dalam *microarray data*. *Data mining* merupakan rangkaian proses untuk mendapatkan sebuah informasi penting yang mulanya tidak diketahui, salah satunya dengan teknik klasifikasi yaitu teknik untuk menggolongkan data yang memiliki karakteristik tertentu yang sama. Dengan teknik tersebut akan dapat menentukan apakah penderita memiliki penyakit kanker [3].

Salah satu penerapan algoritma untuk klasifikasi *microarray data* adalah algoritma *decision tree*, contohnya adalah C4.5. *Decision tree* merupakan metode linier yang mudah diinterpretasi dan termasuk dalam sepuluh algoritma *data mining* yang paling berpengaruh dalam klasifikasi *microarray data*, namun sensitif terhadap data dengan banyak *noise* seperti *microarray data* [4]. *Microarray data* memiliki jumlah *feature* yang sangat banyak (*high dimensional*) dimana tidak semua *feature* tersebut memiliki informasi yang penting (*high noise*) dan jumlah sampel yang sedikit sehingga penerapan teknik klasifikasi pada *microarray data* menjadi sulit karena dapat mempengaruhi nilai akurasi [1]. Untuk mengatasi masalah penerapan teknik klasifikasi pada *microarray data* perlu dilakukan reduksi dimensi sebelum dilakukannya klasifikasi, yakni *feature selection* sebagai alat *pre-processing data*.

Feature selection terdiri dari kategori *univariate* dan *multivariate*. Salah satu metode pada kategori *multivariate* adalah *embedded method*. *Embedded method* memiliki kelebihan yaitu dapat menguji kekuatan prediksi gen [5]. Dalam *embedded method* akan mencari *feature* yang kontribusinya terbaik yang merupakan *feature* yang paling optimal. Pencarian untuk mendapatkan *feature* optimal tersebut dapat menggunakan pendekatan *meta-heuristic*. Salah satu algoritma dari pendekatan *meta-heuristic* adalah *Particle Swarm Optimization (PSO)*. *Feature selection* merupakan masalah yang menangani data diskrit sehingga PSO perlu dimodifikasi menjadi *Binary PSO (BPSO)* [6].

Dari rekomendasi mengenai penelitian selanjutnya dari penelitian sebelumnya [6], BPSO dapat mencapai hasil yang lebih baik jika inisialisasi parameter, nilai iterasi dan penggunaan nilai konstan faktor *weight* dikembangkan lebih lanjut. Pada penelitian ini menerapkan penggunaan nilai konstan faktor *weight*, nilai iterasi generasi yang digunakan sama seperti penelitian sebelumnya dan mengobservasi beberapa nilai inisialisasi parameter untuk BPSO.

Perumusan dan Batasan Masalah

Perumusan masalah dari penelitian ini yaitu bagaimana mengimplementasikan *feature selection* dan klasifikasi terhadap *microarray data*, bagaimana pengaruh diskritisasi terhadap klasifikasi *microarray data*, bagaimana pengaruh *feature selection* terhadap klasifikasi *microarray data*, dan bagaimana performansi yang dihasilkan dari sistem yang dibangun berdasarkan observasi parameter terhadap deteksi kanker.

Batasan dalam penelitian ini yaitu data yang digunakan adalah *ALL-AML Leukimia, Breast Cancer, Colon Tumor, Lung Cancer, dan Ovarian Cancer* yang diambil dari Kent Ridge Biomedical Dataset Repository. *Decision Tree* akan bekerja jika data berbentuk diskrit. Karena data pada *microarray* berbentuk kontinu maka data perlu di diskritisasi. Untuk diskritisasi data pada penelitian ini menggunakan metode *clustering K-Means* dikarenakan merupakan salah satu penerapan diskritisasi data untuk *microarray* yang mengacu pada observasi penelitian sebelumnya [7].

Tujuan

Tujuan dari penelitian ini yaitu mengimplementasikan *feature selection* dan klasifikasi terhadap *microarray data* dengan menggunakan *Binary Particle Swarm Optimization* (BPSO) sebagai *feature selection* dan C4.5 *Decision Tree* sebagai *classifier*, mengetahui pengaruh diskritisasi dan *feature selection* terhadap klasifikasi *microarray data*, dan mengetahui performansi yang dihasilkan dari sistem yang dibangun berdasarkan observasi parameter terhadap deteksi kanker.

Organisasi Penulisan

Setelah bagian Pendahuluan, bagian selanjutnya terdiri dari bagian Studi Literatur, Pembangunan Sistem, Evaluasi dan Kesimpulan. Bagian Studi Literatur memuat teori *microarray*, studi komparatif dan deskripsi metode-metode yang diterapkan pada penelitian ini. Bagian Pembangunan Sistem memuat pembangunan sistem secara umum dan dibagi dalam dua skema secara detail. Bagian Evaluasi memuat hasil pengujian sistem dan analisis hasil pengujian. Bagian Kesimpulan memuat kesimpulan dan saran untuk penelitian selanjutnya.