



UNIVERSITAS INDONESIA

**METODE SELEKSI FITUR BERBASIS PERANKINGAN BOBOT
SECARA MULTI STEP MENGGUNAKAN DEEP LEARNING UNTUK
PENCARIAN BIOMARKER PADA DATA MICROARRAY**

THESIS

MUKHLIS AMIEN

1406522102

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI MAGISTER ILMU KOMPUTER
DEPOK
JUNI 2016**



UNIVERSITAS INDONESIA

**METODE SELEKSI FITUR BERBASIS PERANKINGAN BOBOT
SECARA MULTI STEP MENGGUNAKAN DEEP LEARNING UNTUK
PENCARIAN BIOMARKER PADA DATA MICROARRAY**

THESIS

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Master**

MUKHLIS AMIEN

1406522102

**FAKULTAS ILMU KOMPUTER
PROGRAM STUDI MAGISTER ILMU KOMPUTER
DEPOK
JUNI 2016**

HALAMAN PERSETUJUAN

Judul : Metode Seleksi Fitur Berbasis Perankingan Bobot Secara Multi Step Menggunakan Deep Learning untuk Pencarian Biomarker pada Data Microarray
Nama : Mukhlis Amien
NPM : 1406522102

Laporan Thesis ini telah diperiksa dan disetujui.

XX Januari 2016

Ito Wasito PhD.
Pembimbing Thesis

HALAMAN PERNYATAAN ORISINALITAS

**Thesis ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Mukhlis Amien
NPM : 1406522102
Tanda Tangan :

Tanggal : XX Januari 2016

HALAMAN PENGESAHAN

Thesis ini diajukan oleh :
Nama : Mukhlis Amien
NPM : 1406522102
Program Studi : Magister Ilmu Komputer
Judul Thesis : Metode Seleksi Fitur Berbasis Perankingan Bobot Secara Multi Step Menggunakan Deep Learning untuk Pencarian Biomarker pada Data Microarray

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Master pada Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia.

DEWAN PENGUJI

Pembimbing : Ito Wasito PhD. ()

Penguji : Prof. XXX ()

Penguji : Prof. XXXX ()

Penguji : Prof. XXXXXX ()

@todo

Jangan lupa mengisi nama para penguji.

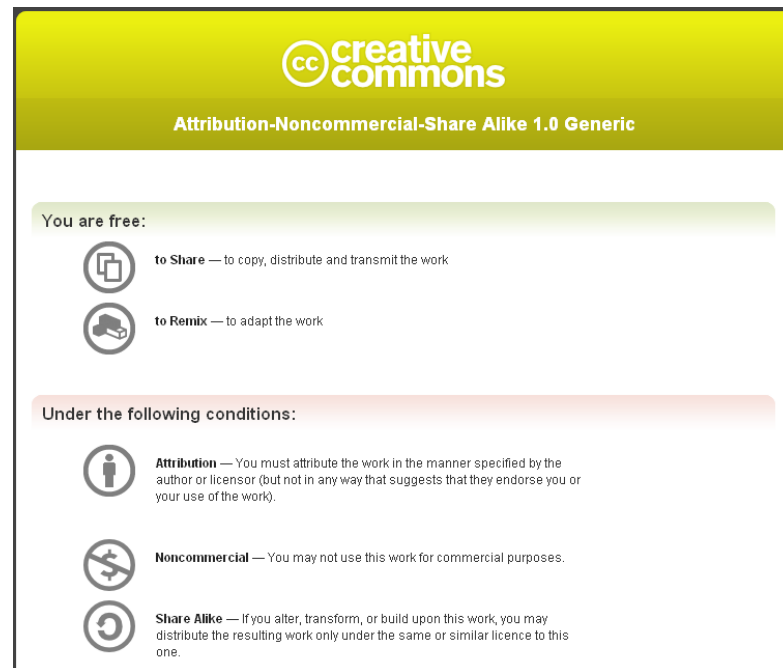
Ditetapkan di : Depok
Tanggal : XX Januari 2016

KATA PENGANTAR

Template ini disediakan untuk orang-orang yang berencana menggunakan \LaTeX untuk membuat dokumen tugas akhirnya. Mengapa \LaTeX ? Ada banyak hal mengapa menggunakan \LaTeX , diantaranya:

1. pertama
 2. kedua
 3. ketiga
1. \LaTeX membuat kita jadi lebih fokus terhadap isi dokumen, bukan tampilan atau halaman.
 2. \LaTeX memudahkan dalam penulisan persamaan matematis.
 3. Adanya otomatis dalam penomoran caption, bab, subbab, subsubbab, referensi, dan rumus.
 4. Adanya automatisasi dalam pembuatan daftar isi, daftar gambar, dan daftar tabel.
 5. Adanya kemudahan dalam memberikan referensi dalam tulisan dengan menggunakan label. Cara ini dapat meminimalkan kesalahan pemberian referensi.

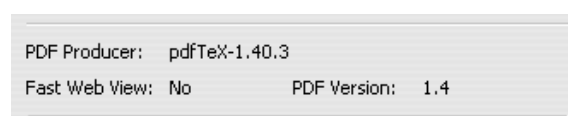
Template ini bebas digunakan dan didistribusikan sesuai dengan aturan *Creative Common License 1.0 Generic*, yang secara sederhana berisi:



Gambar 1: *Creative Common License 1.0 Generic*

Gambar 1 diambil dari http://creativecommons.org/licenses/by-nc-sa/1.0/deed.en_CA. Jika ingin mengetahui lebih lengkap mengenai *Creative Common License 1.0 Generic*, silahkan buka <http://creativecommons.org/licenses/by-nc-sa/1.0/legalcode>. Seluruh dokumen yang dibuat dengan menggunakan template ini sepenuhnya menjadi hak milik pembuat dokumen dan bebas didistribusikan sesuai dengan keperluan masing-masing. Lisensi hanya berlaku jika ada orang yang membuat template baru dengan menggunakan template ini sebagai dasarnya.

Dokumen ini dibuat dengan \LaTeX juga. Untuk meyakinkan Anda, coba lihat properti dari dokumen ini dan Anda akan menemukan bagian seperti Gambar 2. Dokumen ini dimaksudkan untuk memberikan gambaran kepada Anda seperti apa mudahnya menggunakan \LaTeX dan juga memperlihatkan betapa bagus dokumen yang dihasilkan. Seluruh url yang Anda temukan dapat Anda klik. Seluruh referensi yang ada juga dapat diklik. Untuk mengerti template yang disediakan, Anda tetap harus membuka kode \LaTeX dan bermain-main dengannya. Penjelasan dalam PDF ini masih bersifat gambaran dan tidak begitu mendetail, dapat dianggap sebagai pengantar singkat. Jika Anda merasa kesulitan dengan template ini, mungkin ada baiknya Anda belajar sedikit dasar-dasar \LaTeX .



Gambar 2: Dokumen Dibuat dengan PDF \LaTeX

Semoga template ini dapat membantu orang-orang yang ingin mencoba menggunakan \LaTeX . Semoga template ini juga tidak berhenti disini dengan ada kontribusi dari para penggunanya. Kami juga ingin berterima kasih kepada Andreas Febrian, Lia Sadita, Fahrurrozi Rahman, Andre Tampubolon, dan Erik Dominikus atas kontribusinya dalam template ini.

Depok, 30 Desember 2009

Mukhlis Amien

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Mukhlis Amien
NPM : 1406522102
Program Studi : Magister Ilmu Komputer
Fakultas : Ilmu Komputer
Jenis Karya : Thesis

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** atas karya ilmiah saya yang berjudul:

Metode Seleksi Fitur Berbasis Perankingan Bobot Secara Multi Step
Menggunakan Deep Learning untuk Pencarian Biomarker pada Data Microarray

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok
Pada tanggal : XX Januari 2016
Yang menyatakan

(Mukhlis Amien)

ABSTRAK

Nama : Mukhlis Amien
Program Studi : Magister Ilmu Komputer
Judul : Metode Seleksi Fitur Berbasis Perankingan Bobot Secara Multi Step Menggunakan Deep Learning untuk Pencarian Biomarker pada Data Microarray

Data ekspresi gen pada percobaan microarray memiliki ciri khas yaitu jumlah sampel yang sedikit dengan dimensi fitur yang sangat banyak. Algoritma *clustering* dan algoritma *deep learning* merupakan dua algoritma *unsupervised learning*, yang bisa membantu menganalisa data ekspresi gen. Algoritma pemilihan fitur digunakan untuk mendapatkan fitur gen yang paling penting. Dan kemudian akan digunakan algoritma *clustering* untuk mendapatkan struktur cluster dari data ekspresi gen. Pemilihan fitur yang paling informatif dari suatu kasus percobaan microarray, merupakan masalah yang ada pada pemrosesan data ekspresi gen. Sehingga diperlukan eksplorasi lebih lanjut untuk pemilihan fiturnya. Seleksi fitur gen, berdasarkan ranking bobot yang dihasilkan oleh *deep learning*, diharapkan dapat memecahkan masalah seleksi fitur tersebut. Sedangkan metode clustering yang dipakai adalah: Cluster Affinity Search Technique (CAST) , K-Means Clustering dan Hierarchical Clustering. Deep learning adalah metode pembelajaran mesin yang merupakan bagian dari algoritma neural network, metode deep learning yang sering dipakai adalah arsitektur deep believe network (DBN). Pendekatan unsupervised learning pada deep learning dan clustering, diharapkan dapat digunakan untuk membantu peneliti dalam menganalisa data ekspresi gen-nya.

Kata Kunci:

Microarray, ekspresi gen, Algoritma Clustering, feature selection, deep learning, unsupervised learning.

ABSTRACT

Name : Mukhlis Amien
Program : Magister Ilmu Komputer
Title : FEATURE SELECTION METHOD BASED ON MULTI STEP
WEIGHT RANKING USING DEEP LEARNING TO SEARCH
BIOMARKER IN LUNG CANCER MICROARRAY DATA SET

Microarray technology has made possible the profiling of gene expressions of the entire genome in a single hybridization experiment. Since microarray data acquire tens of thousands of gene expression values simultaneously. However, the number of sample usually small. Feature selection and clustering algorithm for microarray data analysis is useful to extract cluster structure and to reduce the high dimensional microarray data and reconstruct to lower dimensional with minimum error possible. Deep learning and clustering is a machine learning method. In this research we will investigate the effectiveness of clustering after or prior dimensionality reduction. The most common deep learning architecture used for dimensionality reduction is deep believe network (DBN) and stacked auto encoder (SAE). Pre training unsupervised learning and greedy layer wise training approach are expected for better dimensionality reduction in microarray datasets compared with other methods.

Keywords:

Microarray, ekspresi gen, Algoritma Clustering, feature selection, deep learning, unsupervised learning.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PERNYATAAN ORISINALITAS	iii
LEMBAR PENGESAHAN	iv
KATA PENGANTAR	v
LEMBAR PERSETUJUAN PUBLIKASI ILMIAH	viii
ABSTRAK	ix
Daftar Isi	xi
Daftar Gambar	xiii
Daftar Tabel	xiv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.2.1 Definisi Permasalahan	2
1.2.2 Batasan Permasalahan	3
1.3 Tujuan	3
1.4 Posisi Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
2 LANDASAN TEORI	5
2.1 Ekspresi Gen	5
2.2 Pemrosesan Data Microarray	6
2.3 Ekstraksi Fitur dan Seleksi Fitur Pada Penelitian Sebelumnya	7
2.4 Deep Learning	7
2.5 Restricted Boltzmann Machine	8

2.6	Deep Believe Network	8
2.7	Multi Layer Perceptron	8
2.8	Logistic Regression	8
3	METODOLOGI PENELITIAN	9
3.1	Gambaran Umum Penelitian	9
3.2	Pengumpulan Data dan Pengolahan Awal	10
3.3	Perancangan Algoritma	10
3.4	Melakukan Testing Arsitektur Deep Learning	10
3.5	Implementasi Metode Perangkingan Bobot Secara Multi Step Untuk Mendapatkan Gen Biomarker	10
3.6	Evaluasi Hasil Perangkingan Dengan Klasifikasi Secara Unsupervised	10
3.7	Perbandingan Hasil Perangkingan Dengan Literatur	10
4	HASIL PENELITIAN DAN PEMBAHASAN	11
4.1	Dataset	11
4.2	Overview Metodologi	11
4.3	Hasil Percobaan RBM Dengan Layer-layer yang Berbeda	11
4.4	Hasil Penerapan Multi Step Ranking Bobot	11
4.5	Hasil Evaluasi Dengan Multi Layer Perceptron	12
4.6	Hasil Evaluasi Dengan Literatur Benferrony Method	12
4.7	Hasil Evaluasi Dengan Literatur Harvard Cancer	12
5	KESIMPULAN DAN SARAN	13
5.1	Kesimpulan	13
5.2	Saran	13
	Daftar Referensi	14
	LAMPIRAN	1
	Lampiran 1	2

DAFTAR GAMBAR

1	<i>Creative Common License 1.0 Generic</i>	vi
2	Dokumen Dibuat dengan PDFLatex	vi
2.1	Ada 23,6% dari keseluruhan fungsi gen yang belum diketahui, sehingga pengetahuan tentang fungsi gen masih belum lengkap. (Häggström, 2014)	5
2.2	Proses Keseluruhan Percobaan Microarray.(Babu, 2004)	6

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Data ekspresi gen pada percobaan *microarray* memiliki ciri khas yaitu dimensi fitur gen yang jauh lebih besar dibandingkan dengan sampel pasien. Masalah tersebut menyebabkan penerapan teknik-teknik pendeteksian penyakit dengan menggunakan data ekspresi gen lebih sulit dan memiliki tantangan tersendiri. Tidak semua data ekspresi gen memiliki signifikansi dalam pendeteksian penyakit. Menurut penelitian Yoon et al. (2006) dan Bandyopadhyay et al. (2014) tidak semua gen yang didapatkan dalam percobaan *microarray* tersebut adalah gen yang informatif, bahkan jumlah ekspresi gen yang informatif untuk kasus yang diinginkan misalnya untuk pengenalan sel kanker, sangat sedikit dibandingkan dengan keseluruhan ekspresi gen yang didapatkan dalam sebuah percobaan (Bandyopadhyay et al., 2014). Data ekspresi gen yang tidak informatif tersebut dapat mengurangi performa proses pengenalan pola secara signifikan pada teknik *machine learning* yang diterapkan. Akan tetapi, beberapa gen yang informatif berpengaruh secara signifikan terhadap pengenalan pola tersebut. Sebagai contoh, untuk mendiagnosa kanker paru-paru, hanya dibutuhkan sekitar 50 gen saja dari 22 ribu gen yang didapatkan dalam percobaan. Gen-gen yang paling informatif ini disebut dengan *Biomarker* (Belinsky, 2004). Sehingga, hanya dengan menggunakan data *Biomarker* yang ditemukan saja, sudah dapat digunakan untuk mengenali penyakit yang diderita oleh pasien.

@todo

motivasi dan kendala pencarian biomarker,
tahap muncul machine learning

Menemukan Biomarker adalah xxx. Pada penelitian ini, akan dibangun sebuah teknik pencarian *Biomarker* dengan metode seleksi fitur gen. Metode yang digunakan merupakan modifikasi dari algoritma seleksi fitur untuk *logistic regression* yang dilakukan oleh Shevade and Keerthi (2003). Akan tetapi, *Logistic regression* memiliki masalah dalam mengeliminasi fitur, dikarenakan koefisien bobot ditempatkan disetiap fitur. Oleh karena itu, pada penelitian ini akan diajukan

teknik pencarian *Biomarker* dengan perankingan bobot pada arsitektur *Deep Learning*. Teknik perankingan bobot ini akan dimodifikasi dan digunakan untuk meranking fitur gennya secara berjenjang atau *multi step*.

@todo

hubungan antara deep learning, bobot dan multistep ranking

Arsitektur *deep learning* yang akan digunakan adalah arsitektur *Deep Belief Network (DBN)*. DBN merupakan jaringan *Restrictive Boltzmann Machine (RBM)* yang disusun secara bertingkat. Dimulai dengan memberikan bobot random diantara dua network, yang dapat dilatih dengan cara meminimalkan perbedaan antara data asli dengan data rekonstruksinya. Gradien didapatkan dengan *chain rule* untuk melakukan penurunan error dengan teknik *Contrastive Divergence (CD)*. Untuk dicari bobot (W) dengan *maximum likelihood learning* secara *greedy* pada tiap layer-nya (Hinton and Salakhutdinov, 2006).

Tahapan berikutnya, fitur *Biomarker* yang telah didapatkan akan digunakan sebagai data untuk penerapan evaluasi apakah *Biomarker* yang ditemukan dengan perankingan tersebut dapat memperbaiki hasil klasifikasi. Untuk mengevaluasi seberapa akurat gen biomarker yang dipilih, dilakukan perbandingan hasil eksperimen dengan literatur.

1.2 Perumusan Masalah

Dikarenakan karakteristik sedikitnya sampel dan besarnya fitur. Serta tidak lengkapnya informasi kita terhadap gen. Apakah pendekatan unsupervised pada deep learning untuk mencari *Biomarker* dengan perankingan bobot secara *multi step* cocok dipakai pada data microarray?

1.2.1 Definisi Permasalahan

@todo

Tuliskan permasalahan yang ingin diselesaikan. Bisa juga berbentuk pertanyaan

1.2.2 Batasan Permasalahan

@todo

pikirkan lagi batasan permasalahan

- Dataset microarray untuk kanker paru-paru yang tersedia secara bebas dengan kode GSE10072
- Data yang digunakan adalah dataset yang sudah dilakukan preprocessing standar dan sudah dinormalisasi.

1.3 Tujuan

Penelitian ini bertujuan untuk membangun metodologi dalam pencarian *Biomarker* Gen yang paling penting untuk percobaan microarray. Dengan menghitung bobot ranking gen secara multi step.

1.4 Posisi Penelitian

@todo

Posisi penelitian Anda jika dilihat secara bersamaan dengan peneliti-peneliti lainnya. Akan lebih baik lagi jika ikut menyertakan diagram yang menjelaskan hubungan dan keterkaitan antar penelitian-penelitian sebelumnya

1.5 Manfaat Penelitian

Mendapatkan framework untuk melakukan perankingan data ekspresi gen sehingga didapatkan gen yang paling informatif menggunakan arsitektur deep learning agar membantu dalam pencarian biomarker pada penelitian data microarray.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Berisi gambaran permasalahan dan metodologi apa yang akan diterapkan

- Bab 2 LANDASAN TEORI
Landasan teori dipakainya metodologi yang akan diterapkan dalam eksperimen
- Bab 3 METODOLOGI PENELITIAN
Penjelasan detail metodologi yang akan diterapkan dalam penelitian
- Bab 4 HASIL PENELITIAN DAN PEMBAHASAN
Pembahasan hasil dari eksperimen
- Bab 5 KESIMPULAN DAN SARAN

@todo

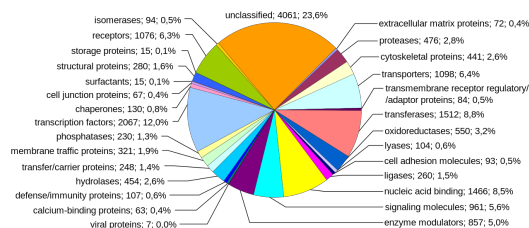
Tambahkan penjelasan singkat mengenai isi masing-masing bab.

BAB 2

LANDASAN TEORI

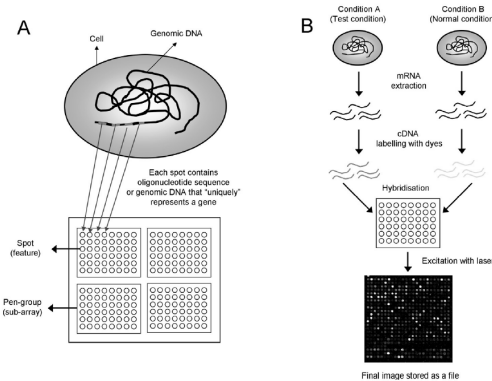
2.1 Ekspresi Gen

Percobaan microarray, mengukur tingkat aktivitas gen di dalam sebuah jaringan sel. Sehingga dapat memberikan informasi berdasarkan aktivitas di dalam jaringan yang bersangkutan. Data ini didapatkan dengan cara mengukur banyaknya mRNA yang diproduksi pada saat proses transkripsi DNA, dimana dapat diukur seberapa aktif atau seberapa berfungsinya gen tersebut dalam sebuah jaringan (Elloumi and Zomaya, 2011). Karena kanker berhubungan dengan berbagai macam aktivitas penyimpangan regulasi pada sel, maka data ekspresi gen pada kanker merefleksikan penyimpangan regulasi tersebut. Untuk menangkap keabnormalan ini, percobaan microarray, dimana dapat mengukur secara simultan dari level ekspresi ratusan bahkan ribuan ekspresi gen dapat digunakan untuk mengidentifikasi kanker. Percobaan microarray sering dipakai untuk membandingkan profil ekspresi gen pada sel yang terkena kanker, dibandingkan dengan sel yang normal pada berbagai macam percobaan. Percobaan microarray digunakan untuk mengidentifikasi ekspresi yang berbeda pada dua percobaan, yang biasanya berupa data tes dan data kontrol.



Gambar 2.1: Ada 23,6% dari keseluruhan fungsi gen yang belum diketahui, sehingga pengetahuan tentang fungsi gen masih belum lengkap. (Hägström, 2014)

Data ekspresi gen yang masih mentah didapatkan dari percobaan di laboratorium menggunakan alat yang dinamakan dengan alat Genchip microarray. Data tersebut kemudian dilakukan pemrosesan awal untuk mendapatkan sebuah matriks ekspresi gen. Matriks ini memiliki data kolom dan baris, dimana kolom berisi data eksperimen, dan baris berisi nilai ekspresi pada tiap-tiap gen (gambar 2.2) [Babu, 2004].



Gambar 2.2: Proses Keseluruhan Percobaan Microarray.(Babu, 2004)

@todo

tambahkan dua gambar yang digambar sendiri

Karena data microarray yang didapatkan dapat mencapai ribuan ekspresi dalam satu waktu secara simultan, maka data ini dapat sangat membantu dalam mengidentifikasi penyakit. Akan tetapi, hasil yang didapat dengan menganalisa beberapa data microarray yang dilakukan oleh dua percobaan yang berbeda tetapi dengan tujuan yang sama, dapat menghasilkan hasil yang sangat berbeda. Salah satu alasannya adalah terbatasnya sampel dan terlalu banyaknya profil ekspresi gen. Sehingga diperlukan metode testing statistik untuk memastikan bahwa data microarray tersebut memiliki tingkat signifikansi yang cukup, dan dipastikan bahwa perbedaan tersebut memang karena eksperimen, bukan karena kerusakan alat atau kesalahan prosedur eksperimen.

2.2 Pemrosesan Data Microarray

Data yang dihasilkan dari alat microarray ini berupa citra yang perlu diproses lebih lanjut. Sebelum data ekspresi gen dapat dianalisa lebih lanjut, perlu dilakukan pemrosesan awal yang berupa (i) perbaikan background, (ii) normalisasi data dan kemudian (iii) penyaringan data.

1. Perbaikan Background

Perbaikan background ini ditujukan untuk menghilangkan titik-titik noise yang tidak berasal dari proses hibridisasi. Metode untuk perbaikan background ini banyak diajukan dalam penelitian. [15]

2. Normalisasi

Tujuan dari normalisasi adalah untuk mengatur bias yang dihasilkan oleh

variasi proses percobaan microarray. Metode normalisasi data microarray ada banyak, dan pada penelitian ini akan digunakan normalisasi standar untuk data microarray.

3. **Penyaringan data** Tidak semua data yang didapat dari percobaan microarray bagus, kadangkala terjadi kesalahan alat dan noise yang diakibatkan oleh alat, oleh karena itu perlu disaring, mana data yang disebabkan oleh proses biologi, dan mana yang disebabkan oleh noise alat.

4. **Missing Value Imputation**

Tidak semua data ekspresi gen dapat kita dapatkan, dikarenakan rumitnya percobaan microarray, kadangkala data tidak kita dapatkan, oleh sebab itu diperlukan metode untuk melakukan pendekatan statistic dalam memberikan perkiraan isi data dalam titik data yang hilang tersebut.

5. **Seleksi Fitur**

Setelah proses diatas, diperlukan teknik untuk menseleksi fitur pada data microarray. Ada banyak metode yang sudah diusulkan oleh para peneliti. Seperti pada table 1 dibawah. Dan pada titik inilah penelitian ini dijalankan. Diharapkan penelitian ini menghasilkan metode reduksi dimensi untuk data microarray.

2.3 Ekstraksi Fitur dan Seleksi Fitur Pada Penelitian Sebelumnya

@todo

tulis tabel perbandingan seleksi fitur

2.4 Deep Learning

@todo

tuliskan secara singkat tentang deep learning

2.5 Restricted Boltzmann Machine

@todo

tuliskan secara detail tentang rbm

2.6 Deep Believe Network

@todo

tuliskan secara detail tentang dbn

2.7 Multi Layer Perceptron

@todo

tuliskan secara detail tentang dbn

2.8 Logistic Regression

@todo

tuliskan secara detail tentang dbn

BAB 3

METODOLOGI PENELITIAN

@todo

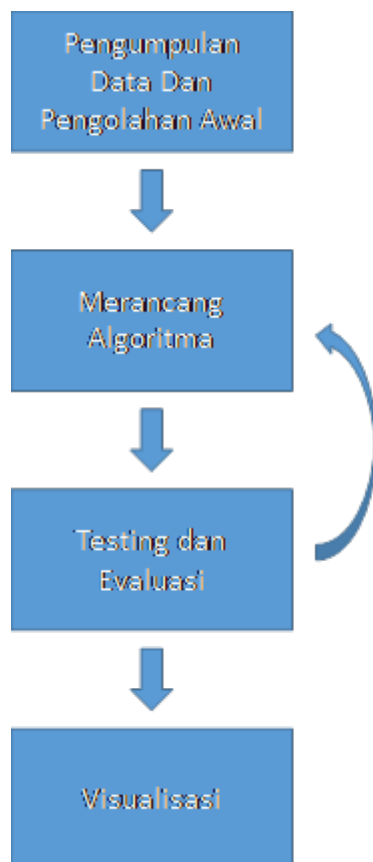
tambahkan kata-kata pengantar bab 3 disini

Penelitian ini dibagi menjadi tiga tahap: (1) Mendapatkan data microarray dan pengolahan awal; (2) Perancangan algoritma; (3) Testing dan kemudian dilanjutkan dengan evaluasi.

@todo

untuk sementara copy paste dari laporan

3.1 Gambaran Umum Penelitian



- 3.2 Pengumpulan Data dan Pengolahan Awal**
- 3.3 Perancangan Algoritma**
- 3.4 Melakukan Testing Arsitektur Deep Learning**
- 3.5 Implementasi Metode Perangkingan Bobot Secara Multi Step Untuk Mendapatkan Gen Biomarker**
- 3.6 Evaluasi Hasil Perangkingan Dengan Klasifikasi Secara Un-supervised**
- 3.7 Perbandingan Hasil Perangkingan Dengan Literatur**

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

@todo

tambahkan kata-kata pengantar bab 4 disini

4.1 Dataset

@todo

Beri keterangan detail dataset yang dipakai

4.2 Overview Metodologi

@todo

buat gambar overview metodologi yang akan dilakukan

4.3 Hasil Percobaan RBM Dengan Layer-layer yang Berbeda

@todo

percobaan 1 percobaan 2 percobaan 3

4.4 Hasil Penerapan Multi Step Ranking Bobot

@todo

top 250 gen pada percobaan 1 2 dan 3, serta diagram venn hasil

4.5 Hasil Evaluasi Dengan Multi Layer Perceptron

@todo

penekanannya pada improvement klasifikasi

4.6 Hasil Evaluasi Dengan Literatur Benferrony Method

@todo

penekanannya pada bahwa hasil percobaan berpotongan xxx persen

4.7 Hasil Evaluasi Dengan Literatur Harvard Cancer

@todo

penekanannya bahwa gen yang ditemukan sangat berkorelasi dengan kanker paru2

BAB 5

KESIMPULAN DAN SARAN

@todo

Tambahkan kesimpulan dan saran terkait dengan pekerjaan yang dilakukan.

5.1 Kesimpulan

5.2 Saran

DAFTAR REFERENSI

- M Mwanadan Babu. Introduction to microarray data analysis. *Computational genomics: Theory and application*, pages 225–249, 2004.
- Supriyo Bandyopadhyay, Saurav Mallik, and Amit Mukhopadhyay. A survey and comparative study of statistical tests for identifying differential expression from microarray data. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 11(1):95–115, 2014.
- Steven A Belinsky. Gene-promoter hypermethylation as a biomarker in lung cancer. *Nature Reviews Cancer*, 4(9):707–717, 2004.
- Mourad Elloumi and Albert Y Zomaya. *Algorithms in computational molecular biology: techniques, approaches and applications*, volume 21. John Wiley & Sons, 2011.
- Mikael Häggström. Diagram of the pathways of human steroidogenesis. *Medicine*, 1:1, 2014.
- Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- Shirish Krishnaj Shevade and S Sathiya Keerthi. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17):2246–2253, 2003.
- Youngmi Yoon, Jongchan Lee, and Sanghyun Park. Building a classifier for integrated microarray datasets through two-stage approach. In *BioInformatics and BioEngineering, 2006. BIBE 2006. Sixth IEEE Symposium on*, pages 94–102. IEEE, 2006.

LAMPIRAN

LAMPIRAN 1

@todo

Membuat todolist apa yang akan dikerjakan untuk thesis

1. belajar copy paste code
3. belajar buat bagan
4. belajar pseudo code
5. kumpulan dalam sebuah tutorial dan link dengan cepat secara offline jika diperlukan
7. export ke odf