Implementasi Mutual Information dan Naive Bayes untuk Klasifikasi Data Microarray

Mohamad Syahrul Mubarok[1], Kurnia C Widiastuti[2], Adiwijaya[3]

[1,2,3]Proram Studi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

[1]msyahrulmubarok@gmail.com, [2]kurniabahariawan@gmail.com, [3]kang.adiwijaya@gmail.com

Abstract. Menurut data dari *World Health Oeganization* (WHO), kanker merupakan salah satu penyakit penyumbang utama kematian didunia. Sekitar 8,2 juta orang meninggal karena kanker. Oleh karena berbahayanya penyakit ini berbagai cara dilakukan untuk melakukan pencegahan maupun pendeteksian secara dini. Namun pendeteksian penyakit kanker sangatlah rentan terjadi berbagai kesalahan yang dilakukan manusia, seperti yang dimuat pada Jurnal kesehatan BMJ, kesalahan medis menyebabkan 251.454 kematian setiap tahunnya di Amerika Serikat. Guna menanggulangi masalah tersebut digunakanlah teknologi *gene expression* dengan bantuan teknologi *microarray*. Masalah muncul, kurang memungkinkan pengolahan data *microarray* karena besarnya dimensi yang dimiliki. Apabila dimensi dikurangi secara tiba-tiba tentu akan merusak informasi yang dimiliki dan akan berakibat data tidak dapat diklasifikasikan. Karenanya dibutuhkan sistem yang mampu mengklasifikasi data *microarray* tanpa menghilangkan informasi penting. Pada penelitian ini, sistem dibangun dengan menggunakan pendekatan *machine learning* yaitu dengan *Naive Bayes*. Untuk mencapai pendekatan ini, dibutuhkan *feature selection* berupa *Mutual Information*. *Feature* tersebut menangani kasus reduksi dimensi, yang mana memisahkan variabel terpenting dari keseluruhan variabel. Untuk mengukur performansi sistem yang dibangun, digunakan *F1-score*. Sistem yang dibangun mampu mengklasifikasi kanker pada data *microarray* dengan rata-rata *F1-score* mencapai 0.89.

1. Pendahuluan

Menurut data dari *World Health Organization* (WHO), kanker dianggap sebagai penyebab utama kematian didunia. Sekitar 8,2 juta orang meninggal karena kanker dan jumlah ini diperkirakan akan terus bertambah pada setiap tahunnya. Dikarenakan berbahayanya penyakit ini maka perlu dilakukan pendeteksian secara dini, cara melakukan pendeteksian dapat dengan melalui pemeriksaan secara fisik, *X-ray* maupun melalui teknologi *gene expression.* Teknologi *gene expression* merupakan salah satu metode pendeteksian kanker dengan bantuan teknologi *microarray* [1].

Teknologi *microarray* adalah salah satu teknologi yang dibangun untuk mempelajari ekspresi dari banyak sifat yang dibawa oleh gen dalam satu waktu [2], alasan ini telah menjadi dasar beberapa penelitian dalam mendeteksi kanker. Dengan teknologi komputasi saat ini pendeteksian kanker dapat dilakukan dalam waktu yang cepat dan dapat mengurangi berbagai kesalahan yang dilakukan oleh manusia, seperti yang dimuat pada jurnal kesehatan BMJ, kesalahan medis menyebabkan 251.454 kematian setiap tahunnya di Amerika Serikat. Namun masalah utama dalam penyelesaian *microarray* adalah mengolah data dalam *microarray,* hal ini dikarenakan dimensi yang dimiliki *microarray* sangat besar.

Masalah pengklasfikasian data *microarray* telah banyak mengundang penelitian melakukan penelitian terkait hal ini, seperti Ramon Diaz [3], Devi Arockia [4], dan Liwei Fan Fan [5]. Dalam [3], dijelaskan secara rinci yang dimaksud dengan *microarray*, detail distribusi data, dan pendekatan untuk memecahkan kasus pengklasifikasian data *microarray*. Banyak pendekatan berbeda yang telah diteliti, seperti [4] yang mengusulkan pendekatan *feature selection* menggunakan *Mutual Information* (MI), pada penelitian tersebut kemampuan MI memiliki performa yang bagus dalam hal mereduksi dimensi. Sedangkan dalam [5], untuk mengklasifikasikan data *microarray*, digunakan *Naive Bayes* (NB), yaitu dengan mengasumsikan ke independenan pada setiap variabel terhadap *class.*
Dalam penelitian Tugas Akhir ini, dihasilkan sistem yang mampu pengklasifikasian data *microarray*. Pendekatan ini diharapkan bisa mendapatkan informasi terpenting dari pola *feature selection* yang ada. Untuk *feature selection* yang digunakan difokuskan dengan menggunakan *Mutual Information*. Dengan demikian, kasus pengklasifikasian data *microarray* dapat ditangani. Sedangkan *mahine learning* yang digunakan adalah *Naive Bayes* (NB), NB digunakan karena mampu mencari pola dan keterhubungan variabel dengan *class* dengan pendekatan probabilitas. Dengan pendekatan ini, sistem yang dibangun ternyata mampu mengklasifikasikan data *microarray* dengan sangat baik.

2. Pendekatan Machine Learning

Fokus utama dari penelitian ini ada dua yaitu, penggunaan metode *mutual information* dan pengklasifikasian *naive bayes*. Untuk tahapan utama sistem dibagi menjadi tiga bagian yaitu *Discretization, Feature Selection,* dan *Learning Naive Bayes*.

2.1. Discretization

Data *input* yang digunakan berasal dari datase. Proses ini merupakan salah satu cara yang digunakan guna mengubah data *continue* menjadi diskrit. Salah satu caranya adalah dengan bantuan metode K Means. Teknik yang digunakan oleh *K means* adalah dengan membagi data menjadi beberapa kelompok.

Untuk menghitung *K means* dengan cara menyebar titik pusat (*centroid)* yang berguna untuk mempresentasikan *cluster* tersebut. Dikarenakan pemberian titik *centroid* diberikan secara acak dan memungkinkan penempatan terjadi secara berdekatan maka perlu dihitung menggunakan

$$\mu_k = \frac{1}{N_k}\sum_{q=1}^{N_k} x_q \qquad (1)$$

Dimana :

μ_k = Titik *centroid* dari *cluster* ke-K

N_k = Banyak data pada *cluster* ke-K

x_q = Data ke-q pada *cluster* ke-K

2.2. Feature Selection

Proses ini merupakan salah satu cara guna menangai permasalahan *microarray* dalam hal menangai besarnya dimensi dalam *microarray. Salah satu caranya* adalah *mutual information.* Teknik yang digunakan oleh *mutual information* adalah dengan memilih gen yang informatif dari keseluruhan ekspresi gen yang asli [6].
Untuk menghitung *mutual information* menggunakan cara histogram dari data, hal ini dikarenakan probabilitas gen yang diperlukan belum diketahui. Langkah-langkah yang perlu dilakukan dalam melakukan perhitungan histogram adalah sebagai berikut:

a. Data set diurutkan secara *ascending* berdasarkan *output*
b. *Output* akan diberikan label (C) yang akan dibagi menjadi dua kelompok dan H(C) akan dihitung menggunakan

$$H(Y) = -\sum_{j=1}^{Ny} P(Yi).\log(P(Yi)) \qquad (2)$$

c. *Input* akan diberikan label (X) akan dibagi menjadi sepuluh kelompok dan H(Y|X) akan dievaluasi menggunakan

$$H(Y|X) = -\sum_{i=1}^{Nx} P(Xi).\sum_{j=1}^{Ny} P(Yj|Xi).\log(P(Yj|Xi)) \quad (3)$$

d. Selanjutnya, *mutual information* dari setiap gen yang berhubungan dengan *output* akan dihitung menggunakan

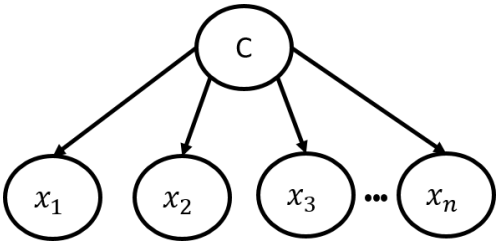
$$I(Y;X) = H(Y) - H(Y|X) \quad (4)$$

Dimana :
Xi = *inputan ke-i* $\{X_1 : 2.5, X_2 : 1.5 \ldots\ldots X_n : 2.25\}$
Yi = *class* ke-i $\{Y_1 : ALL, Y_2 : AML\}$

e. Hasil *Mutual Information* akan diurutkan secara *ascending.* Yang pertama adalah atribut dengan nilai tertinggi dan akan dipilih sebagai gen yang informatif

2.3. Naïve Bayes

Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis, *Bayes Optimal Classifier* [7], menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.
Naïve Bayes, atau dapat dikatakan Teorema Bayes dengan asusmsi keindependenan atribut. Asumsi keindependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latiih dari perkalian kartesius seluruh atribut yang digunakan untuk mengklasifikasikan suatu data [8].



Gambar 1 Contoh Struktur Naive Bayes

Dimana :
C = *class* $\{C_{Leukimia\ ALL-AML} : ALL, AML\}$
x_1 = atribut ke-1 $\{x_1 : 250\}$

Pendekatan *Bayes* pada saat klasifikasi adalah mencari probabilitas tertinggi (V_{map}) dengan memasukan atribut (a_1, a_2, a_3, ..., a_n) seperti Nampak pada persamaan berikut [9]:

$$V_{MAP} = \arg\max P(v_j|a_1,a_2,a_3,\ldots,a_n) \quad (5)$$

Teorema *Bayes* sendiri berawal dari rumus persamaan berikut [9]:

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \quad (6)$$

Dimana P (A | B) artinya peluang A jika diketahui keadaan B. kemudian dari persamaan rumus diatas didapat persamaan seperti berikut [9]:

$$P(B \cap A) = P(B \mid A)\, P(A) \quad (7)$$

Sehingga didapatkan teorema *bayes* seperti persamaan berikut:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \quad (8)$$

Dimana:
A adalah hipotesis data A (class spesifik)
B adalah data dengan class yang belum diketahui
P(A | B) adalah probabilitas hipotesis A berdasarkan kondisi B (posterior | probability)
P(A | B) adalah probalitias B berdasarkan kondisi pada hipotesis A
P(A) adalah probabilitas hipotesis A (prior probability)
P(B) adalah probabilitas dari B

Tahap pertama dari skema ini adalah penentuan struktur NB. Dengan menggunakan MAP *estimation* (*maximum a posteriori*).

Untuk menghitung MAP, diberikan r_i yang menandakan kardinalitas dari X_i, dan q_i merepresentasikan kardinalitas dari *parent set* X_i. Lalu *conditionally probability* $P(X_i|pa(X_i))$ bisa direpresentasikan menjadi $\theta_{ijk} = P(X_i = k|pa(X_i) = j)$, di mana $\theta_{ijk} \in \theta,\ 1 \le i \le n,\ 1 \le j \le q_i$ dan $1 \le k \le r_i$. Kemudian diasumsikan $D = \{D_1, D_2, \dots, D_n\}$ sebagai *dataset* yang diobservasi oleh BN [10].

$$\theta_{ijk} = \frac{N_{ijk} + \alpha_{ijk}}{N_{ij} + \alpha_{ij}} \quad (9)$$

Di mana N_{ijk} adalah jumlah data pada *D* untuk X_i diambil dari nilai *k* dan *parent* $pa(X_i)$ diambil dari nilai *j*. Untuk N_{ij} adalah jumlah data pada *D* untuk setiap $pa(X_i)$ diambil dari nilai *j*. Dengan nilai α_{ijk} dan α_{ij} sebagai berikut.

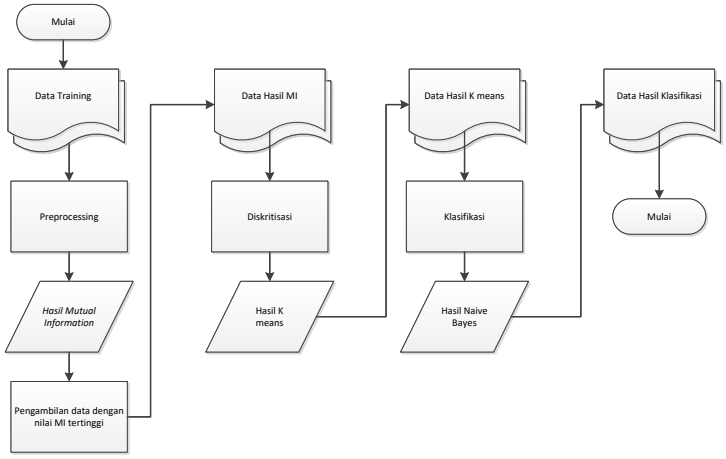
$$\alpha_{ijk} = \frac{\alpha}{r_{ijk}\, q_{ijk}} \quad (10)$$

$$\alpha_{ij} = \frac{\alpha}{r_{ij}\, q_{ij}} \quad (11)$$

Di mana nilai α adalah *hyperparameter* sebagai suatu bilangan yang berfungsi untuk menghindari pembagian nol. Pada sistem ini α = 0.001.

3. Desain Sistem

Alur sistem utama dibagi menjadi dua alur yaitu, alur *training* dan *testing*. Pada alur *training,* inputan data berupa semua data *training.* Tujuan dari proses *training* ini adalah untuk mendapatkan *parameter* dari NB berdasarkan distribusi dari data *training.* Alur dijelaskan pada *flowchart* diagram Gambar 2.
Pada alur *testing,* kurang lebih sama dengan alur *training* yang membedakan adalah proses *learning naive bayes*. Proses tersebut tidak diperlukan, yang diperlukan adalah proses pengambilan data model *naive bayes* yang sebelumnya sudah dilatih untuk kemudian dilakukan inferensi menggunakan *joint probability* terhadap data testing yang ada. Bentuk dari data input *testing* sendiri sama dengan data *input training*.



Gambar 2 Alur Desain Sistem

Performansi sistem diukur pada saat skema *testing* telah dilakukan. Yang artinya menunjukkan performansi model yang dipakai terhadap data baru, dalam hal ini data *testing.* Pengukuran performansinya menggunakan F1 *Score*, dengan formulasi berikut.

$$Precision = \frac{TP}{Total\ Predicted} \quad (12)$$

$$Recall = \frac{TP}{Total\ Target} \quad (13)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

Di mana TP adalah *true postive* yang berarti jumlah model bisa memprediksi kelas dengan benar dari kelas targetnya.

Dengan pengukuran F1 lebih lanjut:

$$F1_{macro} = \frac{\sum_{i=1}^{n} F1_i}{n} \quad (15)$$

$$F1_{micro} = \frac{n}{\sum_{i=1}^{n} \frac{1}{F1_i}} \quad (16)$$

Di mana n adalah jumlah kelas dan $F1_i$ menunjukan skor F1 dari masing-masing kelas berdasarkan order I terhadap n. Pengukuran tersebut diperlukan untuk menilai performansi rata-rata dari setiap kelas yang ada

4. Testing and Analysis

Dataset yang digunakan adalah Kent Ridge Bio-medical [11], di sana tersedia data *training* dan data *testing*

Tabel 1 Distribusi data pada dataset [7]

Data Set	*Number of Instances*	*Number of Genes*	*Number of Classes*
Colon Cancer	62	2000	2
Ovarian Cancer	253	15154	2
Leukima ALL-AML	72	7129	2
Breast Cancer	97	24481	2
Lung Cancer	181	12533	2

Untuk mencapai tujuan penelitian yaitu membangun sistem klasifikasi data *microarray* dengan kombinasi algoritma-algoritma tertentu dengan performansi yang tinggi, maka diperlukan skenario-skenario yang sekiranya bisa menunjang tujuan tersebut.
Skenario pada penelitian ini dibagi menjadi dua yaitu, pertama adalah observasi pengaruh reduksi dimensi dan kedua adalah observasi nilai K pada *K-means*.

4.1. Skenario 1

Pengujian skenario 1 dilakukan dengan bantuan metode *mutual information*. Metode *mutual information* merupakan salah satu metode yang berguna untuk mereduksi dimensi pada data *microarray*. Dengan demikian semua *variable* dapat diobservasi dengan melihat tingkat kepentingan pada setiap *attribute*. Pengujian skenario 1 ini diperlukan guna menangani masalah yang dimiliki oleh data *microarray* sendiri, yaitu berupa besarnya dimensi yang dimiliki oleh data tersebut.
Telah diobservasi untuk *Colon Cancer* dengan K=10, *Ovarian Cancer* dengan K = 8, *Leukimia ALL-AML* dengan K=10, *Breast Cancer* dengan K=10, dan *Lung Cancer* dengan K=12 dengan 6 kombinasi jumlah atribut dengan ketentuan seperti pada Tabel 2.

Tabel 2 Persebaran Nilai MI Pada Setiap Dataset

Kombinasi	*Range* Nilai MI				
	Colon Cancer	***Ovarian Cancer***	***Leukimia ALL-AML***	***Breast Cancer***	***Lung Cancer***
1	0.6544-0.6353	0.6541-0.6394	0.6016-0.6017	0.6849-0.6742	0.6931-0.6931
2	0.6544-0.6274	0.6541-0.6364	0.6016-0.5973	0.6849-0.6703	0.6931-0.6873
3	0.6544-0.6203	0.6541-0.6311	0.6016-0.4587	0.6849-0.6679	0.6931-0.6817
4	0.6544-0.6130	0.6541-0.6278	0.6016-0.3870	0.6849-0.6655	0.6931-0.6814
5	0.6544-0.6055	0.6541-0.6256	0.6016-0.3614	0.6849-0.6633	0.6931-0.6758
6	0.6544-0.5961	0.6541-0.6215	0.6016-0.3407	0.6849-0.6617	0.6931-0.6752

Sehingga berdasarkan persebaran tersebut didapatkan hasil dari skenario-1 terhadap kelima dataset, seperti pada Tabel 3.

Tabel 3 Perbandingan Performansi dari Berbagai Jumlah Atribut

Kombinasi	*Range* Nilai MI				
	Colon Cancer	***Ovarian Cancer***	***Leukimia ALL-AML***	***Breast Cancer***	***Lung Cancer***
	NB	**NB**	**NB**	**NB**	**NB**
1	0.6667	0.9677	0.5882	0.5789	0.9731
2	0.7333	0.9677	0.7647	0.7368	0.9799
3	0.7333	0.9839	0.8823	0.7895	0.9799
4	0.7333	0.9677	0.8823	0.7895	0.9866
5	0.7333	0.9677	0.8823	0.7895	0.9866
6	0.8	0.9516	0.8823	0.7895	0.9866

Dari semua analisis pada setiap dataset didapatkan analisis keseluruhan berupa:

a. Meningkatkan jumlah variabel tidak menjamin dapat meningkatkan hasil performansi, seperti yang di hasilkan oleh dataset *Ovarian Cancer* dan *Leukimia ALL-AML*
b. Terdapat titik masksimal pada beberapa data yang kemudian dapat meningkat terus maupun menurun seperti yang di hasilkan oleh dataset *Breast Cancer, Colon Cancer,* dan *Lung Cancer*

Hal tersebut disebabkan karena perbedaan jumlah variabel yang menjadi ciri khas masing-masing dataset serta persebaran data yang beragam.

4.2. Skenarion2

Pengujian skenario 2 dilakukan dengan bantuan metode *K-means*. Metode *K-means* merupakan salah satu metode yang berguna untuk diskritisasi dengan bantuan *clustering*. Dengan demikian semua nilai pada *variable* yang semula memiliki persebaran data yang besar dapat di minimalisir. Pengujian skenario 2 ini diperlukan guna menangani masalah perhitungan *mutual information*, yaitu perbedaan distribusi antara *variabel* dengan *class*

Tabel 4 Perbandingan Performansi dari Berbagai jumlah variable pada Lung Cancer

Evaluation		K=2	K=8	K=10	K=12	K=16
Colon Cancer	NB	0.6667	0.8	0.7333	0.7333	0.7333
Ovarian Cancer	NB	0.9508	0.9839	0.9839	0.9839	0.5484
Leukimia ALL-AML	NB	0.8823	0.7941	0.8823	0.5882	0.5882
Breast Cancer	NB	0.6842	0.6316	0.7895	0.6316	0.6316
Lung Cancer	NB	0.9799	0.9799	0.9799	0.9866	0.9731
Average	NB	0.8328	0.8379	0.8738	0.7847	0.6949

Dari semua analisis pada setiap dataset didapatkan analisis keseluruhan berupa:

a. Meningkatkan nilai K tidak menjamin dapat meningkatkan hasil akurasi yang akan didapat, seperti pada data *Leukimia.* Hal ini dikarenakan persebaran data pada setiap dataset yang berbeda-beda yang menyebabkan perbedaan nilai K pada setiap dataset.
b. Maksimal rata-rata akurasi berada pada K = 10 yaitu 0.87378. namun hal tersebut tidak berlaku pada data *Leukimia.*
c. Pada dataset *Lung Cancer, Breast Cancer,* dan *Colon Cancer* memiliki titik tertinggi. Dimana pada titik teresebut merupakan titik maksimum yang dapat dihasilkan.

4.3. Perbandingan dan Rata-Rata Hasil Maksimal

Setelah melakukan keseluruhan percobaan maka didapatkan hasil maksimal pada setiap dataset sebagai berikut:

Tabel 5 Perbandingan dan Rata-Rata Hasil Maksimal

Evaluation		Colon Cancer	Ovarian Cancer	Leukimia ALL-AML	Breast Cancer	Lung Cancer	AVG
Max Accurancy	NB	0.8	0.9839	0.8823	0.7895	0.9866	0.8885

Berdasarkan Tabel membuktikan bahwa metode *Mutual Information* sebagai metode *feature selection* dan metode *Naive Bayes* sebagai metode *classifier* mampu mengklasifikasikan data *microarray* dengan rata-rata *accuracy* 0.8885. Hasil tersebut didapatkan berdasarkan analisis pada jumlah variabel dan nilai K yang berbeda pada setiap dataset.

5. Kesimpulan dan Saran

Dalam kasus *microarray* ini hal pertama yang harus diperhatikan adalah data kontinu atau data yang memiliki persebaran data yang sangat luas. Untuk itu diperlukan metode *Clustering* yang dapat

mengatasi masalah ini. Pada penelitian ini digunakan metode *K-Means* untuk *discretization* atau *Clustering.* Jumlah *K, pada K-means* memiliki tren yang berbeda bergantung pada banyaknya random atribut yang digunakan maupun jenis dataset.

Untuk permasalahan seleksi fitur, penelitian ini menggunakan pendekatan metode *Mutual Information.* Dari *feature selection* didapatkan informasi bahwa semakin besar *range* nilai MI yang menyebabkan semakin banyak juga random atribut input yang digunakan tidak selalu menaikan performansi pada setiap dataset. Namun tiap-tiap dataset memiliki batas tersendiri untuk mencapai *F1-score* maksimalnya. Alangkah baiknya jika mengobservasi lebih lanjut metode-metode untuk *discretization* dan *feature selection* agar mendapatkan tren yang lebih baik lagi untuk data-datanya. Karena hal tersebut dapat meningkatkan performansi yang di dapat.

Pendekatan machine learning dengan menggunakan Naive Bayes ternyata mampu mengklafikasi data microarray dengan performasi yang bagus. Dengan nilai performansi rata-rata sebesar 0.88.

References

[1] Nurfalah, Adiyasa; Adiwijaya; Suryani, Arie Ardiyanti, "Cancer Detection Based On Microarray Data Classification Using PCA and Modified Backpropagation," *Far East Journal of Electronics and Communications,* vol. 16, no. 2, pp. 269-281, 2016.

[2] Singh, Rabindra Kumar; Sivabalakrishman, Dr. M.,, "Feature Selection of Gene Expression Data For Cancer Classifiaction: A Review," in *in 2nd International Symposium on Big Data dan Cloud Computing (ISBCC'15)*, Vellore, India, 2015.

[3] Ramon Diaz-Uriarte;Sara Alvarez de Andres, "Methodology Article," BMC Bioinformatics, 06 Januari 2006. [Online]. Available: http://www.biomedcentral.com/1471-2105/7/3. [Accessed 2016].

[4] C, Devi Arockia Vanitha; D, Devaraj; M, Venkatesulu, "Gene Expression Data Classification using Support Vector Machine and Mutual Information-based Gene Selection," in *n Graph Algorithms, High Performance Implementations and Application (ICGHIA2014)*, Tamil Nadu, India, 2015.

[5] fan, Liwei; Poh, Kim-Leng; Zhou, Peng,, "A Sequential Feature Extraction Approach for Naive Bayes Classification of Microarray Data," *Expert System with Applicatnaions,* vol. 36, no. 0957-4174, pp. 9919-9923, 2009.

[6] C, Devi Arockia Vanitha; D, Devaraj; M, Venkatesulu, "Gene Expression Data Classification using Support Vector Machine and Mutual Information-based Gene Selection," in *Graph Algorithms, High Performance Implementations and Application (ICGHIA2014)*, Tamil Nadu, India, 2015.

[7] T. M. Mitchell, Machine Learning, New York: McGraw-Hill, 1997.

[8] Berson, A; J, Smith S., Data Warehousing, Data Mining, & OLAP, New York: McGraw-Hill, 2001.

[9] Almira Syawli; M. Ali Fahmi; Silvia Ari Santhy; Zulkarnaen, "Diagnosa Penyakit Diabetes Mellitus Berbasis Desktop Application," *Jurnal Diabetes Mellitus,* vol. 3, no. 2, pp. 50-62, 2010.

[10] Yun Zhou; Norman Fenton; Martin Neil, "Bayesian Network Approach to Multinomial Parameter Learning Using Data and Expert Judgments," *International Journal od Approximate Reasoning,* vol. 55, pp. 1252-1268, 2014.

[11] Bio-medical, Kent Ridge, "Kent Ridge Bio-medical Dataset," 2003. [Online]. Available: http://datam.i2r.a-star.edu.sg/dataset/krbd. [Accessed 12 2016].

[12] Adiwijaya, Aplikasi Matriks dan Ruang Vektor, Yogyakarta: Graha Ilmu, 2014.

[13] Adiwijaya, Matematika Diskrit dan Aplikasinya, Bandung: Alfabeta, 2016.