

# **PEMBUATAN SISTEM CERDAS UNTUK DETEKSI DINI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA DECISION TREE DAN FUZZY CLUSTERING**

**Titi Mei Tatilah**

Program Studi Teknik Informatika, Universitas Dian Nuswantoro Semarang  
Jl. Nakula I No. 5-11 Semarang, Fax :0243520125, No. Telp : 024 3517261  
E-mail: sekretariat[at]dinus.ac.id

## **ABSTRAK**

Penyakit jantung merupakan salah satu penyakit yang mematikan di Indonesia. Berdasarkan data dari Yayasan Jantung Indonesia, penyakit jantung dan pembuluh darah di Indonesia mencapai 26,8 persen dan semakin mendekati penyebab kematian tertinggi dengan usia penderita yang semakin muda. Deteksi dini pada penyakit jantung sangat diperlukan dalam rangka penurunan angka kematian akibat serangan jantung. Salah satu metode yang dapat membantu dalam ketepatan pengambilan keputusan terkait diagnosa penyakit jantung adalah dengan diimplementasikannya sistem cerdas.

Pada tugas akhir ini, sistem cerdas dibangun melalui kombinasi algoritma decision tree & fuzzy clustering. Algoritma tersebut diterapkan pada aplikasi matlab dengan menggunakan data histori penyakit jantung yang didapatkan dari dataset UCI Machine Learning Repository. Hasil akhir yang menjadi keluaran adalah terbentuknya sistem cerdas yang dapat menjadi tools pelengkap bagi dokter dalam melakukan diagnosa penyakit jantung dengan rendahnya tingkat stadium penyakit ke tinggi. Hasil pengujian dari sistem kinerja menunjukkan bahwa kinerja classifier pohon keputusan memberikan AUC (Area Under ROC) nilai sebesar 0,878. Nilai-nilai ini dikategorikan dalam kategori klasifikasi baik atau dapat didefinisikan bahwa model classifier dapat mengklasifikasikan data tepat.

Kunci : Decision Tree, Fuzzy Clustering, Penyakit Jantung, Sistem Cerdas.

## **ABSTRACT**

*Heart disease is one of the deadly disease in Indonesia. Based on data from the Indonesian Heart Foundation, heart and blood vessel disease in Indonesia reached 26.8 percent and getting closer to the cause of death in patients with an increasingly young age. Early detection of heart disease is needed in order to decrease the death rate from heart attacks. One method that can help in decision making related to diagnostic accuracy of heart disease is the implementation of intelligent systems.*

*At the end of this task, intelligent systems are built through a combination of decision tree and fuzzy clustering algorithms. The algorithm is implemented in matlab applications using the data history of heart disease dataset obtained from UCI Machine Learning Repository. The final result is the creation of intelligent systems that can be complementary tools for doctors in diagnosing heart disease with low levels of disease stage to high. The test result of the performance system shows that the decision tree classifier performance gives the AUC (Area Under ROC) value amount 0.878. These values are categorized in a good classification category or can be defined that the classifier model can classify the data appropriately.*

*Keyword: decision tree, fuzzy clustering, heart disease, intelligent system*

## PENDAHULUAN

Penyakit Jantung merupakan salah satu penyakit tidak menular (PTM) yang menyebabkan kematian tertinggi pada manusia. Berdasarkan data yang dikeluarkan oleh PBB pada tahun 2005, penderita penyakit jantung dan pembuluh darah memiliki prosentase sebesar 29 % atau sebanyak 17,1 juta pasien meninggal dunia setiap tahunnya di seluruh dunia. Berdasarkan data dari Yayasan Jantung Indonesia, Indonesia sendiri memiliki prosentase penderita penyakit jantung sebesar 26,8 %. Jumlah kasus ini mengalami peningkatan setiap tahunnya dengan usia penderita yang semakin muda. Deteksi dini pada penyakit ini sebenarnya dapat dilakukan melalui diagnosa yang tepat dari seorang dokter (Antara, 2011). Akan tetapi permasalahan yang dihadapi adalah tidak semua dokter di Indonesia memiliki keahlian yang sama dalam melakukan diagnosa pada penyakit jantung. Selain itu, tidak semua rumah sakit di Indonesia memiliki fasilitas yang memadai untuk melakukan pemeriksaan penyakit jantung secara menyeluruh. Berdasarkan permasalahan yang telah dikemukakan di atas, maka diperlukan tools terotomotasi berupa sistem cerdas yang berisi kumpulan knowledge dari para ahli yang dalam hal

ini adalah dokter spesialis penyakit jantung dalam melakukan prediksi penyakit jantung. Selain itu diharapkan sistem cerdas yang dibangun dapat menjadi alternatif atau pelengkap dalam melakukan diagnosa penyakit jantung. Pada penelitian sebelumnya telah dilakukan uji coba pada sistem cerdas Intelligent Heart Disease Prediction System (IHDPS) dengan menggunakan tiga algoritma yaitu algoritma decision tree, naive bayes & artificial neural network (ANN).

Hasil dari penelitian tersebut menunjukkan nilai akurasi masing-masing 89 %, 86,53 % dan 85 %. Algoritma decision tree menunjukkan prediksi dengan tingkat akurasi yang paling tinggi (Soni, Ansari, Shama, & Soni, 2011). Penelitian lain yang masih berkaitan dengan penelitian sebelumnya yakni prediksi penyakit jantung menggunakan algoritma klasifikasi yang diproses melalui aplikasi WEKA machine learning. Hasil penelitian tersebut menunjukkan hasil bahwa algoritma decision tree memiliki tingkat akurasi yang paling tinggi yakni 99.2 % dibandingkan dengan algoritma naive bayes dan classification via clustering, masing-masing dengan prosentase akurasi sebesar 96.5 % dan 88.3 % (M, E, & N.CH.S.N, 2010). Selain penelitian

menggunakan algoritma decision tree, pada penelitian terkait telah dipergunakan kombinasi algoritma decision tree dan fuzzy c-means untuk prediksi dan klasifikasi breast cancer yang diderita oleh pasien. Algoritma fuzzy cmeans merupakan algoritma yang telah umum dan sudah banyak diterima banyak pakar dalam melakukan teknik klasterisasi. Dari penelitian tersebut didapatkan hasil dengan akurasi yang cukup tinggi yakni sebesar 92,3 % (Shanthi & Murali Bhaskaran, 2011). Berdasarkan beberapa penelitian yang telah dikemukakan sebelumnya, pada tugas akhir ini dipergunakan dua kombinasi algoritma tersebut. Algoritma decision tree digunakan untuk menentukan diagnosa penyakit jantung, sedangkan algoritma fuzzy clustering ( Fuzzy C-Means ) digunakan untuk klasterisasi stadium penyakit jantung yang diderita oleh pasien. Sebagai hasil akhir dari tugas akhir ini, diharapkan akan dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan penelitian yang telah dilakukan sebelumnya.

A. Kelebihan dari metode pohon keputusan adalah:

1. Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global,

dapat diubah menjadi lebih simpel dan spesifik.

2. Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
3. Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama.
4. Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
5. Dalam analisis multivariat, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut.
6. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

B. Kelebihan dari teori logika fuzzy adalah:

1. Kemampuan dalam proses penalaran secara bahasa (linguistic reasoning). Sehingga dalam perancangannya tidak memerlukan persamaan matematik dari objek yang akan dikendalikan.

## HASIL DAN PEMBAHASAN

### 4.2 UJI COBA DAN ANALISIS HASIL

Bagian ini menjelaskan tentang uji coba aplikasi menggunakan data testing, untuk kemudian dilakukan analisa untuk menentukan tingkat validitas hasil dari aplikasi yang dihasilkan.

#### 4.2.1 Lingkungan Uji Coba

Lingkungan uji coba yang digunakan dalam melakukan implementasi dari aplikasi ini terdiri dari dua yakni lingkungan uji coba perangkat keras dan lingkungan uji coba perangkat lunak. Berikut ini akan merupakan penjelasan lengkap tentang masing-masing lingkungan uji coba tersebut :

**Tabel 4.5 Lingkungan Uji Coba Perangkat Keras**

Perangkat Keras	Spesifikasi
Processor	Intel® Core™ 2 Duo CPU E7500 @ 2.93
Disk Drive	5GB

Lingkungan uji coba perangkat lunak terdiri dari:

**Tabel 4.6 Lingkungan Uji Coba Perangkat Lunak**

Nama Perangkat Lunak	Fungsi
Windows	Sistem operasi
Matlab R2009a	Compiler kode program
WEKA 3.4	Generalisasi / Formulasi Dari Sekumpulan Data Sampling

#### 4.2.2 Uji Coba

Bagian ini menjelaskan tentang uji coba yang dilakukan pada aplikasi, data yang digunakan untuk uji coba beserta analisa hasil yang didapatkan dari aplikasi. Uji coba yang dilakukan terdiri dari dua bagian yakni uji coba validasi dan uji verifikasi.

##### 4.2.2.1 Tampilan Halaman Login

##### Aplikasi IHDPS

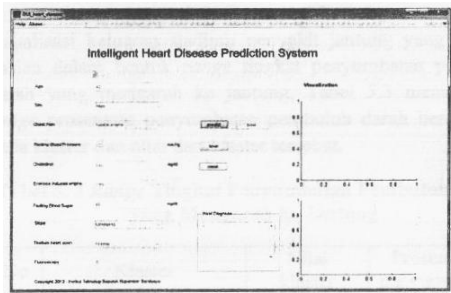


**Gambar 4.12 Halaman login**

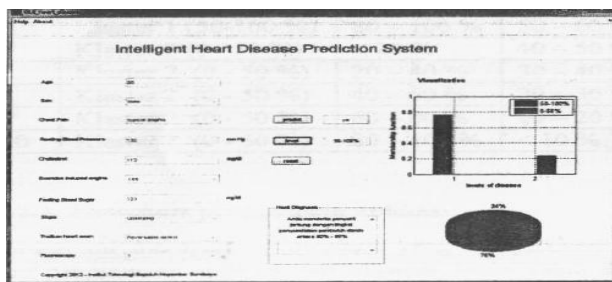
##### Aplikasi IHDPS

Gambar 4.12 merupakan tampilan awal aplikasi Intelligent Heart Disease Prediction System (IHDPS). Pada halaman login ini pengguna memasukkan username dan password. Kemudian apabila masukan username yang dimasukkan pengguna sama dengan password, maka pengguna akan diarahkan pada halaman utama dari aplikasi IHDPS.

Gambar 4.13 menunjukkan tampilan aplikasi IHDPDS yang menunjukkan hasil prediksi 'tidak' atau dapat dinyatakan bahwa pasien tidak sedang menderita penyakit jantung.



**Gambar 4.13 Tampilan aplikasi yang menunjukkan prediksi normal**

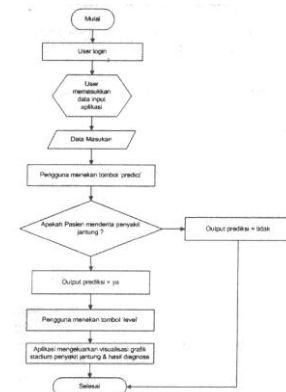


**Gambar 4.14 Visualisasi level & hasil diagnosa penyakit jantung**

Gambar 4.14 menunjukkan hasil untuk prediksi jantung, seperti yang terlihat pada gambar di atas, aplikasi dapat memberikan visualisasi keluaran stadium penyakit jantung yang diderita pasien dalam bentuk range tingkat penyumbatan pembuluh darah yang mengarah ke jantung. Tabel 5.3 menunjukkan range persentase penyumbatan pembuluh darah berdasarkan pada klaster dan nilai dari klaster tersebut.

#### 4.2.2.2 Flowchart penggunaan aplikasi

Berikut ini merupakan flowchart yang menjelaskan tentang cara penggunaan aplikasi dari halaman login yakni dari awal pengguna memasukkan data username dan password sampai dengan hasil akhir prediksi penyakit jantung.



**Gambar 4.15 Flowchart penggunaan aplikasi**

Setelah semua data terisi, berikutnya pengguna menekan tombol predict, prediksi dari aplikasi akan menunjukkan dua nilai yakni prediksi 'ya' dan 'tidak'. Apabila pasien diprediksi 'ya' (atau pasien menderita penyakit jantung), maka langkah selanjutnya adalah melihat stadium penyakit jantung yang diderita pasien dengan cara menekan tombol 'level'. Pada langkah terakhir ini, pengguna dapat melihat visualisasi stadium penyakit yang diderita pasien dalam bentuk graft dan hasil diagnosa yang diberikan kepada pasien. Kemudian, apabila pengguna ingin memasukkan kembali data inputan ke dalam aplikasi, maka pengguna dapat menekan tombol 'reset' untuk menghilangkan isian data yang dimasukkan oleh pengguna sebelumnya.

#### 4.2.2.3 Data Uji Coba

Data yang digunakan untuk proses uji coba berjumlah 30 data (10 %). Data uji ini merupakan data yang telah dipisahkan dari dataset heart disease dari total data sebanyak 297 record.

#### 4.2.2.4 Uji validasi

##### 4.2.2.4.1 Uji validasi klasifikasi

Validasi pada klasifikasi dilakukan dengan penghitungan precision, recall, tingkat akurasi (accuracy) dan penghitungan menggunakan AUC (Area Under ROC). Tabel di bawah ini menunjukkan hasil uji coba menggunakan data testing :

**Tabel 4.9 Confusion Matriks Menggunakan Data Testing**

ACTUAL CLASS	EDICTED CLASS		
		ss = yes	ss = no
	ss = yes		
	ss = no		

Berdasarkan tabel confusion matriks di atas, maka pengukuran precision & recall dari keluaran yang dihasilkan adalah sebagai berikut :

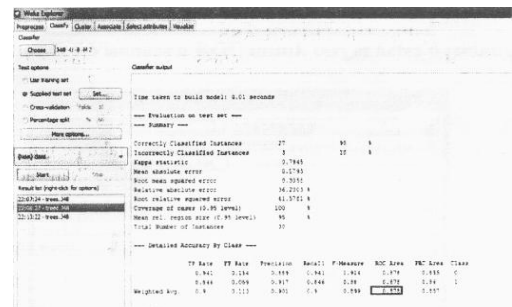
$$Precision = \frac{TP}{TP + FP} \times 100 = \frac{11}{11 + 1} \times 100 = 91.67\%$$

$$Recall = \frac{TP}{TP + FN} \times 100 = \frac{11}{11 + 2} \times 100 = 84.61\%$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$= \frac{11 + 16}{11 + 16 + 1 + 2} \times 100 = 90\%$$

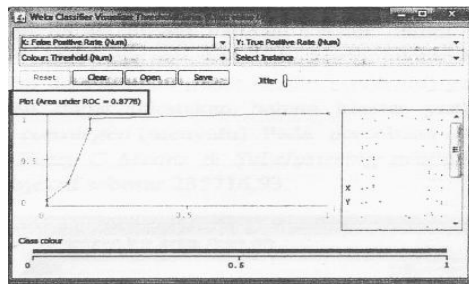
Untuk uji validasi terakhir yang digunakan pada bagian klasifikasi dilakukan pengukuran dengan menggunakan nilai AUC (Area Under the ROC Curve). Berdasarkan pengujian yang dilakukan dengan menggunakan weka machine learning dengan uji test menggunakan metode cross-validation, model classifier memiliki nilai AUC rata-rata untuk kedua kelas sebesar 0.878.



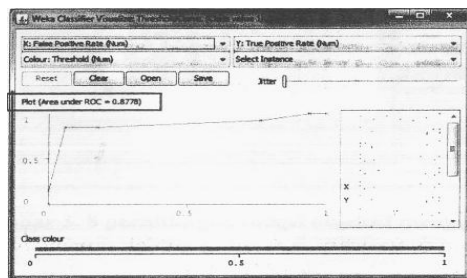
**Gambar 4.16 Nilai AUC rata-rata untuk model classifier**

Berdasarkan standar penilaian model classifier menggunakan AUC (Range penilaian dapat dilihat pada Tabel 3.2), model classifier yang dihasilkan pada sistem cerdas ini dapat digolongkan ke dalam kategori good classification, yaitu dapat menggolongkan instance data secara baik.

Gambar 4.17 dan gambar 4.18 merupakan tampilan kurva ROC pada masing-masing atribut kelas (kelas 0 & 1). Seperti yang tampak pada gambar di bawah, pada bagian atas dari grafik juga ditampilkan nilai dari AUC (Area under ROC Curve).



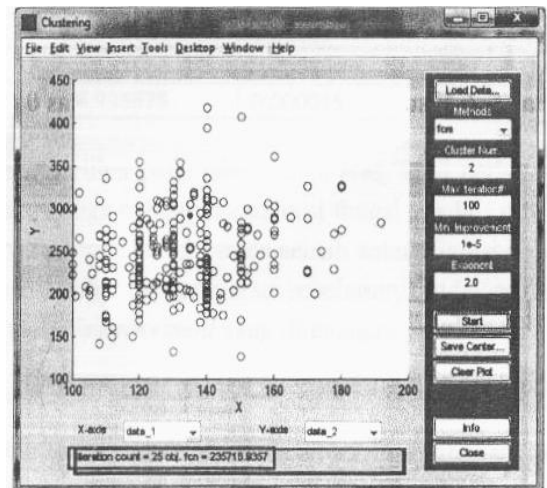
**Gambar 4.17 Visualisasi ROC untuk target kelas 0 (tidak)**



**Gambar 4.18 Visualisasi ROC untuk target kelas 1 (ya)**

#### 4.2.2.4.2 Uji validasi klasterisasi

Uji validasi pada klasterisasi dilakukan dengan membandingkan nilai fungsi objektif (J) antara hasil perhitungan manual dengan hasil dari fungsi FCM yang ada pada matlab. Perbandingan dilakukan pada nilai fungsi objektif karena fungsi objektif merupakan nilai SSE yang digunakan dalam proses validasi klaster dalam FCM. Nilai fungsi objektif yang semakin kecil menunjukkan bahwa jarak antara tiap data dengan pusat klaster (centroid) yang semakin kecil atau dapat dikatakan bahwa klaster yang terbentuk semakin convergen (menyatu). Pada percobaan menggunakan toolbox Fuzzy C Means & Subclustering menunjukkan nilai fungsi objektif sebesar 235716,93.



**Gambar 4.19 Perhitungan fungsi objektif menggunakan toolbox fuzzy c means & subclustering**

Sedangkan melalui perhitungan manual dengan menggunakan excel didapatkan hasil fungsi objektif sebagai berikut :

Seperti yang terlihat pada tabel 5.10 di atas, hasil perhitungan secara manual juga menunjukkan hasil fungsi objektif minimal berhenti pada iterasi ke 25 karena selisih antara fungsi objektif pada iterasi tersebut dengan iterasi sebelumnya tidak melebihi batas minimum improvement yang ditentukan yakni Ke-5.

#### 4.2.2.5. Verifikasi

Verifikasi dilakukan dengan cara membandingkan hasil keluaran dari aplikasi yang dibuat dengan hasil perhitungan manual. Pada aplikasi Intelligent Heart Disease Prediction System (IHDPDS) dilakukan pengukuran terhadap nilai fungsi objektif. Dari hasil keluaran aplikasi, didapatkan nilai fungsi objektif yang sama dengan hasil perhitungan manual, yakni sebagai berikut :





diderita oleh pasien juga didasarkan pada hal yang sama.

Oleh karena itu, pemberian label pada klastering dapat dibagi menjadi dua yakni klaster 1 adalah klaster "50-100%" sedangkan klaster dua adalah klaster "0-50%".

## SIMPULAN DAN SARAN

Berdasarkan percobaan yang telah dilakukan dalam tugas akhir ini, maka dapat disimpulkan bahwa:

1. Algoritma *decision tree* memberikan perbandingan nilai *true positif* dengan *false positif* sebesar 0.878 dalam bentuk nilai *AUC*. Nilai *AUC* ini termasuk dalam kategori *good classification*. Jadi dapat disimpulkan bahwa model *classifier* yang dibuat telah dapat mengklasifikasikan data dengan baik.
2. Algoritma *Fuzzy C Means* menggolongkan data ke dalam dua klaster dengan nilai *SSE* (Fungsi Objektif) sebesar 235716,93567. Penggolongkan ke dalam dua klaster dapat dikategorikan telah optimal karna nilai *SSE* yang dibentuk merupakan nilai fungsi objektif yang minimal dan tidak melebihi batas eror (*minimum improvement*) yang telah ditetapkan sebesar 1-e5.

3. Model *classifier* yang dibuat menunjukkan tingkat efisiensi waktu yang lebih baik dibandingkan dengan penelitian sebelumnya yakni hanya membutuhkan waktu 0.01 *second* dalam melakukan prediksi penyakit jantung.

## DAFTAR PUSTAKA

- [1] (2007). Retrieved Februari 22, 2012, from UCI Machine Learning repository:<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- [2] Analytics: *Decision Tree Induction*. (n.d.). Retrieved June 27, 2012, from meta data offer knowledge:<http://gautam.lis.illinois.edu/monkmiddleware/public/analytics/decisiontree.html>
- [3] (2011). Gangguan Kardiovaskuler. In F. K. ANDALAS, *PENUNTUN SKILLS LAB* (pp. 25-32). PADANG.
- [4] Antara. (2011, September 15). *surya*. Retrieved Februari 15, 2012, from <http://www.surya.co.id/2011/09/15/268persen-kasus-penyakit-jantung-di-indonesia>
- [5] Cardiology Consultants of Philadelphia, Inc. (2008). *Cardiology Consultants of Philadelphia, P.C.* Retrieved maret 7, 2012, from <http://www.ccpdocs.com/faq.aspx>
- [6] Cimino, M. G., Frosini, G., Lazzerini, B., & Marcelloni, F. (2005). On the Noise Distance in Robust. *World Academy of Science, Engineering and Technology* .
- [7] Clinic, M. (2007, Februari 23). Retrieved Maret 18, 2012, from Coronary angiography: A look inside your heart's blood vessels: <http://www.cgi.cnn.com/HEALTH/library/coronaryangiography/HB00048.html>
- [8] Elish, K. O., & Elish, M. O. (2008). Predicting defect-prone software modules using support. *The Journal of Systems and Software* , 653.
- [9] Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Springer. Han, J., & Kamber, M. (2006 ). In *Data mining: concepts and techniques* (pp. 47-97). Sanfrancisco: Morgan Kaufmann.
- [10] Höppner, F. ( 2011, September 29 ). *Ostfalia University of Applied Sciences*. Retrieved Februari 23, 2012, from <http://public.fh-wolfenbuettel.de/~hoeppnef/clustering.html>
- [11] Limitra, B., Kastubi, P. R., & Kurniawan, R. (2009). *ANALISIS DAN PERANCANGAN MODEL PREDIKTIF AIRLINE NO SHOW RATE BERBASIS PASSENGER NAME RECORD PADA PT. METRO BATAVIA*.

- [12] Luthfi, E. T. (2007). FUZZY C-MEANS UNTUK CLUSTERING DATA. *Seminar Nasional Teknologi 2007*, (pp. 1-7). Yogyakarta.
- [13] M, A., E, A., & N.CH.S.N, I. (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. *International Journal of Engineering Science and Technology*, 1-7.
- [14] National Center for Biotechnology Information. (2009, November). *U.S. National Library of Medicine*, 81 *National Institute of health*. Retrieved Maret 18, 2012, from <http://www.ncbi.nlm.nih.gov/pubmed/19917381>
- [15] National Heart Lung and Blood Institute. (2011, Juni 1). Retrieved Maret 7, 2012, from National Heart Lung and Blood Institute, people sciencz helath:<http://www.nhlbi.nih.gov/health/healthtopics/topics/angina/>
- [16] Nightingale Publishing, Inc. (2006, Juni 22). Retrieved Maret 7, 2012, from easydiagnosis: <http://easydiagnosis.com/secondopinions/newsletter26.html>
- [17] Politecnico di milano. (2010). *Politecnico di milano*. Retrieved Maret 3, 2012, from A Tutorial on Clustering Algorithms: [http://home.dei.polimi.it/matteucc/Clustering/tutorial\\_html/index.html](http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/index.html)
- [18] Shanthi, S., & Murali Bhaskaran, V. (2011). Intuitionistic Fuzzy C-Means and Decision Tree Approach for. *European Journal of Scientific Research*, 345-351.
- [19] Soni, J., Ansari, U., Shama, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview. *International Journal of Computer Applications*, 17.
- [20] stack exchange inc. (2011, 3 12). stackoverflow. Retrieved Maret 4, 2012, from <http://stackoverflow.com/questions/1960430/decision-tree-in-matlab>
- [21] Tape, T. G. (n.d.). The Area Under an ROC Curve. Retrieved July 11, 2012, from Interpreting Diagnostic Tests: <http://gim.unmc.edu/dxtests/roc3.htm>
- [22] The Gale Group, Inc. (2008). Retrieved Maret 18, 2012, from Gale Encyclopedia of Medicine: <http://medicaldictionary.thefreedictionary.com/Thallium+Heart+S can>
- [23] The MathWorks, Inc. (1994-2012). The MathWorks, Inc. Retrieved Februari 20 2012, from <http://www.mathworks.com/products/matlab/>
- [24] The MathWorks, Inc. (1984-2012). Retrieved Maret 4, 2012, from <http://www.mathworks.com/help/toolbox/fuzzy/fp310.html>
- [25] Y.L.M, D. (2011, April 10). Retrieved Maret 18, 2012, from Diagnosing hearts: <http://thestar.com.my/health/story.asp?file=/2011/4/10/health/8436288&sec=health>
- [26] Yanowitz, F. G. (n.d.). Retrieved Maret 11, 2012, from The Alan E. Lindsay ECG Learning Center inCyberspace: [http://library.med.utah.edu/kw/ecg/ecg\\_outline/Lessons8/index.html](http://library.med.utah.edu/kw/ecg/ecg_outline/Lessons8/index.html)