

ANALISIS SOFTWARE HYBRID UNTUK DETEKSI GEJALA PENYAKIT

Ilham

Universitas Islam Negeri Sunan Ampel Surabaya,
Jl.Ahmad Yani 117 Sby
ilham@uinsby.ac.id

Abstrak

Penelitian ini dilakukan untuk melihat kecenderungan penyakit yang ditimbulkan akibat pola hidup yang tidak sehat pada masyarakat tertinggal dan pesisir desa-desa sekitar kabupaten Gresik dan Tuban dengan menggunakan algoritma hybrid melalui konstruksi struktur Bayesian Network. Masalah yang akan diselesaikan pada penelitian ini adalah belum ada system yang mampu mendeteksi hubungan antara perilaku hidup tidak sehat dengan penyakit yang di timbulkan. Model struktur ini belum pernah diaplikasikan secara langsung di lapangan untuk mendeteksi suatu kejadian sebab akibat misalnya jika suatu perilaku itu tidak sehat maka akan timbul suatu penyakit. Pengaplikasian model ini perlu dilakukan dengan suatu studi lapangan untuk mengetahui dan membuktikan manfaat yang sebenarnya dari konsep konstruksi struktur hybrid bayesian network. Tujuan dari penelitian ini adalah menghasilkan model perangkat lunak yang mampu mendeteksi dini kecenderungan resiko penyakit masyarakat desa tertinggal dan pesisir yang memiliki pola hidup tidak sehat dalam bentuk konstruksi struktur dan menghasilkan probabilitas nilai dengan cendrung penyakit. Hasil perbandingan antara struktur asal dengan struktur hasil uji coba menunjukan tingkat kecocokan untuk data uji lengkap selisih 10% dengan struktur aslinya, dan tingkat kecocokan untuk data uji tidak lengkap sebesar lebih dari 20% tergantung jumlah missing value nya. Validitas bahwa orang merokok akan mengalami kecenderungan punya penyakit Tuberculosis, Bronkitis atau Lung cancer melalui uji system adalah 80% sampai 90%.

Kata Kunci: Model, Bayesian Network, Pesisir, dan Algoritma Hybrid

Abstract

This study was conducted to see the trend of diseases caused by unhealthy lifestyles in underserved communities and coastal villages around Gresik and Tuban by using a hybrid algorithm through the Bayesian Network structure construction. The purpose of this study is to produce a software model that is able to detect early disease risk propensity lagging rural and coastal communities who have unhealthy lifestyle in the form of construction of structures and produces a probability value with cendrung disease. The validity of that people have a tendency to smoke will experience disease Tuberculosis, Bronchitis or Lung Cancer through the test system is 80% to 90%.

Keywords: Models, Algorithms Hybrid, Lifestyle, and Structural

1. PENDAHULUAN

Bayesian Network merupakan teknik atau metode yang digunakan untuk membangun model klasifikasi data dalam bentuk graf yang dapat merepresentasikan tingkat ketergantungan dari suatu variabel acak dengan menggunakan struktur graf [SIM06]. Model struktur ini belum pernah diaplikasikan secara langsung di lapangan untuk mendeteksi suatu kejadian sebab akibat misalnya jika suatu perilaku itu tidak sehat maka akan timbul suatu penyakit. Pengaplikasian model ini perlu dilakukan dengan suatu studi lapangan

untuk mengetahui dan membuktikan manfaat yang sebenarnya dari konsep konstruksi struktur hybrid bayesian network.

Konsep Bayesian Network mempunyai konstruksi struktur yang cerdas dalam membentuk suatu model yang menggambarkan atau merepresentasikan hubungan sebab akibat (causalitas) untuk suatu kondisi atau perilaku. Sehingga sistem ini perlu dilakukan implementasi dan uji coba lapangan secara langsung. Bayesian Network memiliki dua algoritma yang dapat bekerja pada basisdata lengkap dan tidak lengkap.

Algoritma ini terdiri dari algoritma CB (Constrain Base) dan algoritma BC (Bound and Collapse).

Algoritma CB (Constrain Base) adalah algoritma yang dapat bekerja pada basis data lengkap, sedangkan algoritma BC (Bound and Collapse) adalah algoritma yang dapat bekerja pada basis data yang tidak lengkap (missing value). Masalah yang akan diselesaikan pada penelitian ini adalah membuktikan bagaimana algoritma hybrid ini mampu untuk mengelola data-data perilaku pola hidup tidak sehat masyarakat desa tertinggal dan pesisir agar dapat dideteksi dan diketahui lebih dini akan penyakit ditimbulkannya dengan membentuk konstruksi struktur dan menampilkan nilai-nilai probabilitas kecenderungan.

Tujuan dari penelitian ini adalah menghasilkan model perangkat lunak yang mampu mendeteksi dini kecenderungan resiko penyakit masyarakat desa tertinggal dan pesisir yang memiliki pola hidup tidak sehat dalam bentuk konstruksi struktur dan menghasilkan probabilitas nilai dengan cendrungan penyakit yang ditimbulkan. Implementasi perangkat lunak algoritma hybrid ini juga dapat mengetahui dan mempelajari fungsi dan cara kerja algoritma secara benar sesuai dengan kebutuhan pengaplikasian sehingga dapat diterapkan dalam uji deteksi dini resiko penyakit atau gangguan terhadap kesehatan masyarakat kedepannya sehingga dapat diantisipasi lebih dini.

2. METODE PENELITIAN

2.1 Metode Konstruksi Struktur Bayesian Network

Ada dua pendekatan yang digunakan untuk membangun framework Bayesian Network [CHE01] yaitu metode search and scoring dan metode dependency analysis. Masing-masing metoda ini memandang Bayesian Network dari sudut yang berbeda. Pada metoda search and scoring, Bayesian Network dipandang sebagai sebuah struktur yang merepresentasikan JPD dari variabel- variabel, sedangkan pada metoda dependency analysis, Bayesian Network dipandang sebagai sebuah struktur yang merepresentasikan sekumpulan kebebasan kondisional di antara node-node.

Berikut ini diberikan penjelasan lebih rinci mengenai kedua pendekatan tersebut [CHE97[b]:

1. Metode Search and Scoring (Scoring Based). Metode ini melakukan pengembangan struktur dengan mencari sebuah struktur yang paling cocok dengan data. Dalam metode ini, permasalahan digambarkan sebagai pencarian

struktur graf (model) yang baik dari data. Model dibangkitkan dari kombinasi node (variabel) yang ada. Pencarian tersebut dilakukan dengan search method. Setiap model dievaluasi oleh sebuah scoring function.

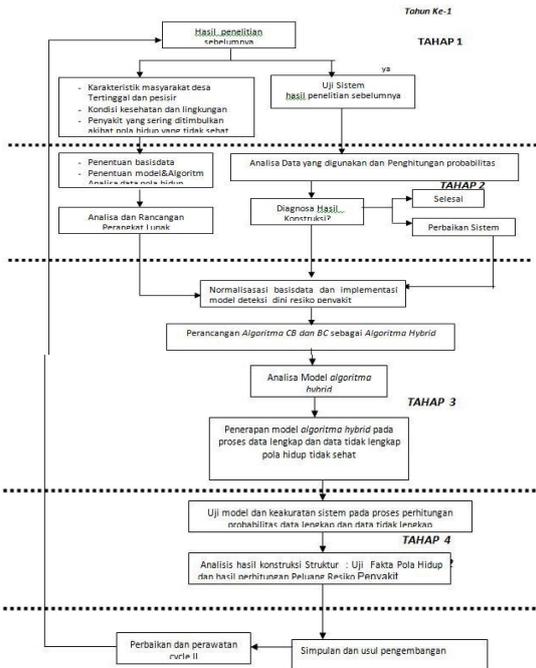
2. Metode Dependency Analysis (Constraint Based) Pada metode ini, permasalahan digambarkan sebagai pencarian kebebasan (ketidakbergantungan) dari data yang kemudian dipergunakan untuk (meng-infer) sebuah struktur. Hubungan kebebasan tersebut data diukur dengan menggunakan salah satu dari beberapa jenis Conditional Independence test (CI test) dan hubungan tersebut digunakan sebagai batasan untuk membangun framework Bayesian Network.

2.2 Algoritma Hybrid Bayesian Network

Algoritma Hybrid merupakan pengembangan dari Algoritma CB dan BC. Algoritma CB dikembangkan oleh Singh & Voltorta adalah sebagai suatu algoritma yang mengkombinasikan metode analisis dependency dan metode search and scoring merupakan salah satu algoritma yang tidak terlalu bergantung pada CI test dan tidak membutuhkan node ordering tetapi akan membangkitkan sendiri node ordering [SIN95].

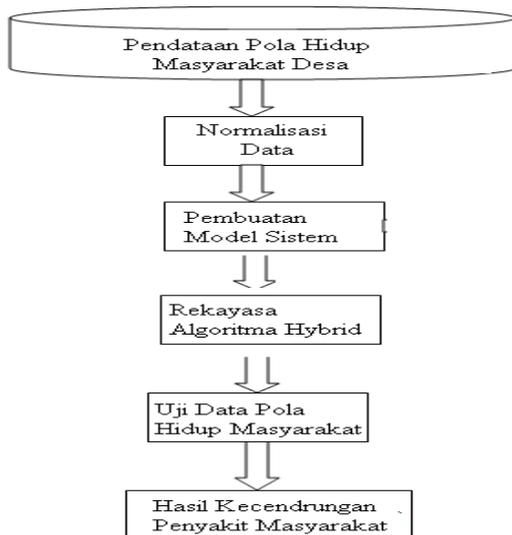
Algoritma CB perlu dikembangkan agar dapat bekerja pada data yang tidak lengkap untuk melakukan membangun struktur Bayesian Network. Pengembangan algoritma ini diberi nama dengan Algoritma Hybrid yaitu dengan mengkombinasikan Algoritma CB dengan Algoritma BC. Algoritma CB akan membangkitkan node ordering kemudian node ordering tersebut akan digunakan sebagai input oleh Algoritma BC agar membangun framework Bayesian Network dari data yang lengkap dan tidak lengkap dapat dilakukan. Di bawah ini adalah susunan lengkap dari Algoritma CB dan BC (algoritma Hybrid) yang akan dikembangkan dalam bentuk perangkat lunak sebagai media alat uji data yang akan diteliti

Adapun rancangan proses penelitian ini terlihat pada Gambar 2.1 sebagai berikut :



Gambar 2.1. Rancangan Proses Sistem

Langkah-langkah detail dalam rekayasa algoritma hybrid dan implementasi yang dilakukan dalam penelitian ini terlihat pada Gambar 2.2 sebagai berikut:



Gambar 2.2. Langkah Implementasi Sistem

Penjelasan langkah algoritma Hybrid pada Gambar 2.2 adalah di mulai dari mengumpulkan data melalui survey dan observasi pola hidup masyarakat desa tertinggal di kabupaten Gresik dan Tuban. Kemudian melakukan normalisasi

terhadap data yang di kumpulkan, penerapan model, rekayasa algoritma hybrid(CB dan BC), Uji data (implementasi system dilapangan) dan hasil resiko penyakit.

Bayesian Network untuk kasus Visit to Asia yang akan dibuktikan tingkat kecocokan atau kesamaannya dengan konstruksi struktur pada Gambar 2.3.

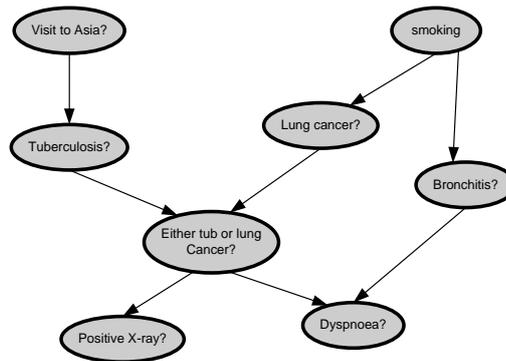
2.3 Hasil Uji Data Pola Hidup Tidak Sehat

Setelah menyelesaikan tahap implementasi, maka selanjutnya tahapan yang akan dilakukan adalah pengujian terhadap perangkat lunak yang telah dihasilkan. Struktur yang dihasilkan untuk kasus uji pertama ini akan dibandingkan dengan struktur asal Bayesian Network Tabel 1.1

Tabel 1.1 Data Hasil Uji Kecocokkan Struktur
Tabel 4.10 Hasil Uji Porsentasi Kecocokkan Struktur

Data Sampel Missing (%)	Tingkat Kecocokan Hasil Konstruksi Struktur			
	Konstruksi Struktur Bayesian Network	Konstruksi Struktur Hasil	Porsentasi (%) Kecocokkan Konstruksi Struktur Asal	Porsentasi (%) Kecocokkan Konstruksi Struktur Hasil
Data	VisitAsia - Tuberculosis	VisitAsia -> Tuberculosis	10	10
VisitAsia	Smoking - LungCancer	Smoking -> LungCancer	10	10
Lengkap	Smoking - Bronchitis	Tuberculosis -> TubOrLung	10	10
	TubOrLung - X-Ray	LungCancer -> TubOrLung	10	10
	Bronchitis - Dyspnoea	Smoking -> Bronchitis	10	10
	Tuberculosis - TubOrLung	TubOrLung -> X-Ray	10	10
	LungCancer - TubOrLung	Bronchitis -> Dyspnoea	10	10
	TubOrLung - Dyspnoea		10	-
			80	70
Data	VisitAsia - Tuberculosis	VisitAsia -> Tuberculosis	10	10
VisitAsia	Smoking - LungCancer	Smoking -> LungCancer	10	10
Missing	Smoking - Bronchitis	LungCancer -> TubOrLung	10	10
Value 50	TubOrLung - X-Ray	X-Ray -> TubOrLung	10	-
	Bronchitis - Dyspnoea	Tuberculosis -> TubOrLung	10	10
	Tuberculosis - TubOrLung	Smoking -> Bronchitis	10	10
	LungCancer - TubOrLung	Bronchitis -> Dyspnoea	10	10
	TubOrLung - Dyspnoea		10	-
		80	60	

Hasil perbandingan antara struktur asal dengan struktur hasil uji coba menunjukkan tingkat kecocokan untuk data uji lengkap selisih 10% dengan struktur aslinya, dan tingkat kecocokan untuk data uji tidak lengkap sebesar lebih dari 20% tergantung jumlah missing value nya. Validitas bahwa merokok akan mengalami kecenderungan punya penyakit Tuberculosis, Bronkitis atau Lung cancer melalui uji system adalah 80% sampai 90%.



Gambar 2.3 Bayesian Network Kasus VisittoAsia

3. KESIMPULAN

Dari hasil uji coba salah satu kasus pola hidup tidak sehat terdapat perubahan pada struktur yang dihasilkan. Perubahan struktur ini dapat terjadi akibat pengaruh data yang digunakan. Data yang digunakan dalam pengujian ini jumlahnya sangat terbatas dan tidak ada jaminan bahwa data tersebut sudah mencukupi sebagai data uji untuk menghasilkan struktur yang baik. Pada satu sisi hasil uji itu menunjukkan perubahan struktur tidak ditemukan walaupun ada data yang hilang sebanyak 50%, hal ini disebabkan karena sampel data yang dimiliki oleh data tersebut berjumlah lebih besar dari dua kasus yang lainnya. Sehingga meskipun data yang hilang adalah 50% namun sudah cukup membuktikan bahwa kecenderungan hasil struktur pola hidup dengan resiko penyakitnya mendekati kecocokan dengan nilai 80% dengan struktur aslinya.

DAFTAR PUSTAKA

Journal Article

- [1] [SIT06] Sitohang, B., & Saptawati, P. (2006). Improvement of CB & BC Algorithm (CB* Algorithm) for Learning Structure of Bayesian Networks as Classifier in Data Mining . Sekolah Teknik Elektro dan Informatika, ITB.J.ICT Vol.1,No.1,2007,29-41
- [2] [PUS05] Sandhyaduhita, P., I. (2005). Algoritma CB: Algoritma yang Dibangun dengan Dua Pendekatan untuk Konstruksi Struktur Bayesian Network dalam Data Mining . Program Studi Teknik Informatika, STEI, ITB.
- [3] [Sel08] Selvia Lorena Br Ginting(2008). Studi algoritma CB Dalam Data Mining untuk Konstruksi Struktur Bayesian Network dari Basis Data Incomplete, Konferensi Nasional Sistem dan Informatika 2008; Bali, November 15, 2008,KNS,108-038
- [4] [CHE97[a]] Cheng, J., Bell, D., & Liu, W. (1997). An Algorithm for Bayesian Belief Networks Construction from Data . Proceeding of Ai & STAT '97 (pp.83-90). Ft. Lauderdale, Florida.
- [5] [CHE97[b]] Cheng, J., Bell, D., & Liu, W. (1998). Learning Bayesian Networks from Data : An Efficient Approach Based on Information Theory. Faculty of Informatics, University of Ulster, U.K.
- [6] [SEB97] Sebastiani, P., & Ramona, M. (1997). Bayesian Inference with Missing Data Using Bound and Collapse . Report KMi-TR-58, Knowledge Media Institute, The Open University.
- [7] [SIN95] Singh, M., & Valtorta, M. (1995). Construction of Bayesian Network Structures from Data: a Brief Survey and an Efficient Algorithm . Dept. of Computer Science, University of South Carolina,Columbia, USA.
- [8] [SIM06] Simanjuntak, H. (2006). Pengembangan Algoritma CB untuk Konstruksi Struktur Bayesian Network dari Data Tidak Lengkap . Program Studi Teknik Informatika, STEI, ITB.
- [9] [MAH05] Maharani, H. (2005). Konstruksi Struktur Bayesian Network dalam Data Mining untuk Basis Data Incomplete dengan Metode Bound and Collapse . Program Studi Teknik Informatika, STEI, ITB.
- [10] [NEA04] Neapolitan, R., E. (2004). Learning Bayesian Networks. USA : Pearson Pentice Hall.
- [11][CHE01] Cheng, Jie, et al. (2001). Learning Bayesian Network from data : An Information-Theory Based Approach . Department of Computing Science, University of Alberta., Faculty of Informatics, University of Ulster, Toronto,Canada.