STRUKTUR PEMBELAJARAN JARINGAN BAYESIAN MENGGUNAKAN ALGORITMA TREE AUGMENTED NETWORK (TAN) UNTUK DIAGNOSIS HIPOTIROID

Nur Salman*[1], Mustikasari[2]
[1]STMIK Dipanegara Makassar, [2]UIN Alauddin Makassar
[1,2]Program Studi Teknik Informatika
e-mail: *[1]nursalman.halim@dipanegara.ac.id, [2]mustikasari@uin-alauddin.ac.id

Abstrak

Diagnosis dini suatu penyakit menghindari kehilangan waktu perawatan yang berharga. Diperlukan metode yang cepat dan akurat untuk mendeteksi fitur secara otomatis dan merampingkan proses diagnosis yang dapat membantu mendeteksi beberapa jenis gejala dari penyakit hipotiroid. Bayesian Network telah digunakan untuk diagnosis inferensi beberapa penyakit karena kemampuannya untuk menyandikan hubungan antara satu set variabel di bawah ketidakpastian. Namun, menemukan pembelajaran struktur terbaik DAG adalah masalah NP-Hard. Penelitian ini menunjukkan bahwa Bayesian Network mengeksplorasi ruang pencarian dari algoritma tree augmented network dan memperlihatkan hubungan antar fitur dalam sebuah struktur. Terakhir, algoritma yang diusulkan dibandingkan dengan algoritma pembelajaran struktur lainnya berdasarkan akurasi klasifikasi serta waktu konstruksi. Hasil eksperimen menunjukkan bahwa algoritma BN-TAN mengungguli algoritma yang lainnya untuk akurasi dan cukup baik untuk waktu konstruksi pada dataset hipotiroid.

Kata kunci — Bayesian Network, Hipotiroid, Struktur Pembelajaran, TAN

Abstract

Early diagnosis of disease avoids losing valuable treatment time. A fast and accurate method is needed to detect features automatically and streamline the diagnosis process which can help detect certain types of symptoms of hypothyroid disease. Bayesian Network has been used for diagnosis of inference of several diseases as of its ability to encode relationships between a set of variables under uncertainty. However, finding the best structure learning of the DAG is an NP-Hard problem. This research shows that the Bayesian Network explores the search space of the tree augmented network algorithm and shows the relationship between features in a structure. Finally, the proposed algorithm compared with other structure learning algorithms based on classification accuracy as well as construction time. Experimental results showed that the BN-TAN algorithm outperforms the competed algorithm for the accuracy of the hypothyroid dataset, and good enough for construction time.

Keywords— Bayesian Network, Hypothyroidism, Learning Structure, TAN

1. PENDAHULUAN

Diagnosis penyakit secara dini sangat penting untuk perencanaan perawatan kesehatan jangka panjang pada pasien. Keakuratan hasil klasifikasi mengenai penyakit tertentu akan berdampak pada meningkatnya peluang penanganan dan pengobatan penyakit yang lebih presisi serta menciptakan kualitas hidup yang lebih baik di masa depan.

Hormon tiroid mempengaruhi secara langsung atau tidak langsung hampir semua sel di dalam tubuh manusia. Diantaranya adalah metabolisme, kalorigenik, system kardiovaskuler, pertumbuhan dan sistem syaraf. Penyakit hipotiroid yang terkait dengan hormon tiroid adalah salah satu jenis penyakit yang terjadi ketika kelenjar tiroid menjadi kurang aktif[1].

Diagnosa untuk penyakit hipotiroid merupakan sebuah permasalahan klasifikasi dalam bidang data mining. Sebuah model dibangun berdasarkan kasus pengklasifikasian dan kontrol sebelumnya, dengan catatan telah diketemukan model penyaringan dan atribut acuan yang menjadi dasar pengklasifikasiannya. Klasifikasi adalah salah satu metode data mining yang paling populer digunakan di sektor kesehatan. Metode ini membagi sampel data sesuai kelas target. Teknik klasifikasi menklasifikasi kelas target untuk setiap poin data. Pendekatan klasifikasi faktor risiko yang dikaitkan pada pasien dilakukan dengan menganalisis pola penyakit. Model ini lebih lanjut dapat digunakan untuk kasus diagnosa baru atau klasifikasi lainnya ke dalam kelas-kelas yang ada.

Penelitian pada paper ini yaitu klasifikasi penyakit hipotiroid menggunakan struktur pembelajaran *Bayesian network* dengan algoritma TAN, diharapkan dapat mengadopsi keutamaan hubungan sebab akibat setiap atribut yang sangat alami pada *Bayesian Network*, sehingga mampu menerapkan sifat ketergantungan antar atribut pada proses nyata untuk selanjutnya dapat menklasifikasi penyakit hipotiroid. Efisiensi dalam hal performa (akurasi) klasifikasi pada algoritma TAN yang diklaim lebih tinggi daripada simple metode seperti *Naïve Bayes* yang telah diakui keakuratannya jauh sebelumnya[2], dan ditambah dengan kemampuan spesifik BN dalam menggambarkan hubungan dibawah ketidakpastian, diharapkan dapat memberi jawaban tentang bagaimana menklasifikasi penyakit hipotiroid secara cepat dan akurat dengan memanfaatkan struktur pembelajaran *Bayesian Network* menggunakan algoritma TAN (BN-TAN).

2. METODE PENELITIAN

Metode penelitian yang dilakukan pada penelitian ini merupakan metode eksperimental. Metode ini merupakan salah satu metode kuantitatif dengan melakukan percobaan untuk melihat hasil dan menyelidiki kemungkinan sebab akibat.

2.1 Data

Data yang digunakan pada penelitian ini, diambil dari UCI *Machine Learning Repository*[3]. Data terdiri atas 3772 data (*instance*). Dataset ini tersusun atas 30 atribut, yaitu ada 7 data kontinu (1 *integer* dan 6 *real*), 23 data diskrit, dan kurang lebih 5,4 % *missing* data (diabaikan) dari total keseluruhan data (yaitu 6064 *instance*).

Data *outlier* yang didapatkan dihilangkan dari data yang ada agar tidak mengganggu. Dalam dataset, satu data teridentifikasi sebagai outlier pada penelitian ini (yaitu data no. 1365) sehingga tidak dimasukkan ke dalam eksperimen, di mana untuk data umur bernilai 445 (adalah sangat tidak realistis). Bentuk data terdiri atas beberapa data kontinu, maka terlebih dahulu diubah menjadi diskrit dengan teknik diskritisasi bin sederhana menjadi 10 bin. Untuk penyakit hipotiroid sendiri telah dikelompokkan ke 4 kelas, yaitu kelas *negative* dan tiga kelas utama lainnya yaitu *composated*, *primarly* dan *secondary* hipotiroid.

2. 2 Bayesian Networks (BN) Dan Struktur Pembelajaran

Bayesian Networks atau BN adalah grafik acylic terarah yang mewakili sambungan distribusi probabilitas satu set variabel random (*node*). Seperangkat parameter numerik, biasanya distribusi probabilitas bersyarat yang diambil dari struktur grafik. *Bayesian Networks* sebagai struktur probabilistik memfaktorkan *Joint Probability Distribution* (JPD) dari satu set variabel random menggunakan data observational.

Bayesian Network menerapkan struktur grafis untuk menyandikan hubungan antara serangkaian variabel dalam kondisi masalah. Struktur grafis ini, atau lebih tepatnya DAG, G =

(V, E), di mana node dalam V = {X1, X2,. . . , Xn} mewakili variabel acak dari masalah yang modelkan, dan topologi grafik (busur dalam E ⊆ V × V) mengkodekan hubungan ketergantungan bersyarat (dalam) antara variabel (dengan cara ada atau tidak adanya koneksi langsung antara pasangan variabel).

Metode ini yang menggunakan prinsip dan teorema Bayes dengan asumsi bahwa atribut input dapat saling terhubung/terikat. Node-node DAG mewakili variabel diskrit dan busur menunjukkan ketergantungan independen dari variabel. Setiap node DAG memiliki tabel probabilitas kondisional atau CPT yang menyajikan probabilitas setiap keadaan simpul sesuai dengan kombinasi setiap keadaan induk.

Proses konstruksi BN dapat dipisahkan dalam dua langkah utama: pembelajaran parameter dan pembelajaran struktur. Karena BN adalah model statistik, dan pembelajaran parameter adalah metode belajar dalam statistik, maka dari itu juga dapat digunakan dalam jaringan Bayesian. Pada dasarnya, pembelajaran parameter adalah perhitungan probabilitas bersyarat untuk struktur BN yang diperoleh dan parameter dalam BN adalah nilai probabilistik dalam *Conditional Probability Table* (CPT). Tujuan utama dari pembelajaran struktur ini adalah menemukan struktur terbaik untuk BN yang kompatibel dengan dataset yang ada. Harapannya bahwa hasil dapat optimal dari sudut kompleksitas dan waktu konstruksi. Algoritma inferensi digunakan untuk jaringan Bayesian sebagai pengklasifikasi, kita cukup menghitung Argmax P(y | x) menggunakan distribusi P (U) yang diwakili oleh jaringan Bayesian[4].

$$P\ (y|\mathbf{x}) = \frac{P(U)}{P(\mathbf{x})}$$

$$= \prod_{u \in U} P(u|p(au)) \quad (1)$$

Algoritma pembelajaran struktur jaringan Bayesian dapat dikelompokkan ke dalam dua kategori termasuk algoritma berbasis kendala dan algoritma berbasis skor. Algoritma berbasis kendala mempelajari struktur jaringan dengan menganalisis hubungan probabilistik yang disyaratkan oleh properti Markov dari jaringan Bayesian dengan kondisional tes independensi dan kemudian membangun grafik yang memenuhi pernyataan pemisahan yang sesuai.

Model yang dihasilkan sering ditafsirkan sebagai model kausal bahkan ketika dipelajari dari data pengamatan. Algoritma berbasis kendala semuanya didasarkan pada algoritma sebab akibat induktif (IC), yang menyediakan kerangka kerja teoritis untuk mempelajari model struktur sebab akibat. Tetapi algoritma berbasis skor memberikan skor kepada setiap kandidat jaringan Bayesian dan mencoba memaksimalkannya dengan beberapa algoritma pencarian heuristik. Pencarian dengan algoritma *greedy* adalah pilihan umum, tetapi hampir semua jenis prosedur pencarian dapat digunakan. Algoritma berbasis skor, di sisi lain hanyalah aplikasi dari berbagai algoritma pencarian heuristik.

2. 2. 1 Tree Augmented Network (TAN)

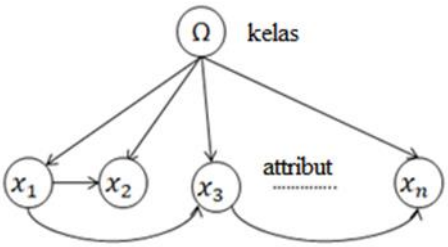
Mempertimbangkan bahwa fitur dapat dikorelasikan dan asumsi independensi dari pengklasifikasi *Naïve Bayes* tidak realistis untuk klasifikasi yang memiliki hubungan antar atribut, Friedman et al. memperkenalkan pohon pengklasifikasi *Tree Augmented Bayes* (TAN)[5]. Metode ini mendorong pengusulan representasi kompromi, menggabungkan beberapa kemampuan jaringan Bayesian untuk mewakili ketergantungan, dengan kesederhanaan Naive Bayes. Representasi ini didefinisikan oleh kondisi berikut:

a. Setiap atribut memiliki atribut kelas sebagai orangtua.
b. Atribut mungkin memiliki satu atribut lainnya sebagai orangtua

Sebagaimana yang dijelaskan sebelumnya bahwa metode ini didasarkan pada struktur jaringan Naïve Bayes dimana variabel kelas adalah induk dari setiap atribut, Oleh karena itu,

probabilitas posterior P(C|X_1,…,X_n) memperhitungkan semua atribut. Selain itu, tepi (busur) di antara atribut diizinkan untuk menangkap korelasi diantara mereka. Setiap atribut mungkin memiliki paling banyak satu atribut lainnya sebagai induk tambahan yang berarti ada busur dalam grafik dari fitur X_1 ke fitur X_n. Ini menyiratkan bahwa dua atribut ini X_1 dan X_n tidak independen berdasarkan label kelas.

Graph Struktur *Bayesian Network* dengan *Tree Augmented Network* (TAN) ditunjukkan pada Gambar 1. Pohon ini terbentuk dengan menghitung berat maksimum pohon spanning (*spanning tree*). Sebagaimana yang jabarkan pada Persamaan (1) dan strukturnya pada BN dengan Algoritma TAN yang selanjutnya akan disingkat BN-TAN pada makalah ini.



Gambar 1 Struktur dari Tree Augmented Network (TAN)

2. 2. 2 Hill-Climbing

Metode ini menambah dan menghapus busur tanpa urutan tetap dari variabel. Pembalikan busur dipertimbangkan ketika menentukan langkah selanjutnya yang akan dibuat. Algoritma ini berupaya untuk memaksimalkan atau meminimalkan fungsi target, di mana elemen yang dipertimbangkan adalah variabel diskrit atau kontinu.

2. 2. 3 Tabu Search

Metode tabu adalah melakukan *hill-climbing* (pendakian bukit) hingga mencapai optimal lokal. Kemudian ia melangkah ke kandidat yang paling tidak buruk di lingkungan itu. Namun, itu tidak mempertimbangkan poin di lingkungan yang baru saja dikunjungi di langkah tabu-list terakhir. Ketika ditemukan solusi lanjutan untuk dikunjungi, beberapa elemen solusi (move) dianggap tabu. Sehingga tidak dapat digunakan untuk konstruksi berikutnya.

Adapun fitur dasar algoritma ini adalah:

a. Selalu pindah ke solusi lingkungan terbaik yang tersedia, bahkan jika itu lebih buruk daripada solusi saat ini.
b. Menyimpan daftar poin solusi yang harus dihindari (*forbidden*) atau daftar fitur pemindahan yang tidak diizinkan yang disebut sebagai daftar tabu.
c. Perbarui daftar tabu berdasarkan pada beberapa struktur memori (memori jangka pendek). Hapus bergerak tabu setelah periode waktu tertentu yang telah berlalu.
d. Izinkan pengecualian dari daftar tabu (*aspiration conditions*). Perluas area pencarian, ubah jangka waktu atau ukuran daftar tabu.

2. 2. 4 K2

Algoritma K2 adalah algoritma berbasis skor yang menemukan struktur terbaik dengan menggunakan proses iteratif antara semua topologi yang mungkin. Ini adalah metode *hill-climbing* yang menambahkan busur dengan urutan variabel tetap. Sebelum pelaksanaan fungsi penilaian, variabel harus dipesan dan jumlah yang tetap dan terbatas, harus dipertimbangkan untuk *parent* dari setiap *node*. Biasanya, dalam aplikasi spesifik, para ahli menentukan urutan *node* dan jumlah *parent* untuk masing-masing simpul.

Algoritma-algoritma tersebut umum digunakan dalam struktur pembelajaran Bayesian[4] pembelajarannya menetapkan skor untuk topologi yang mungkin dan mencoba untuk memaksimalkannya menggunakan fungsi penilaian metrik[6].

2. 3 Validasi

Cross Validation merupakan teknik validasi yang digunakan pada penelitian ini, teknik ini secara random membagi data k bagian dan setiap bagian akan dilakukan proses klasifikasi. Digunakan untuk menemukan nilai tingkat kesalahan keseluruhan pada data latih. Secara umum, pengujian k dilakukan 10 kali untuk memperkirakan akurasi estimasi atau disebut validasi silang 10 kali lipat[7].

$$Akurasi = \frac{(\text{True Positive(TP)} + \text{True Negative(TN)})}{(\text{TP} + \text{FN} + \text{TP} + \text{FP})} \times 100\% \quad (2)$$

Demikian pula dalam penelitian ini. Setiap percobaan akan menggunakan satu tes data dan k-1 bagian akan menjadi data pelatihan (data yang digunakan sebagai pembelajaran), dan data pengujian (data yang belum digunakan sebagai pembelajaran) akan ditukar dengan satu bagian data pelatihan sehingga untuk setiap percobaan akan mendapatkan pengujian data yang berbeda. Rumus akurasi dijabarkan seperti pada Persamaan (2).

Cross-validasi berbeda dari metrik penilaian lokal dalam hal kualitas. Struktur jaringan seringkali tidak dapat didekomposisi dalam skor masing-masing *node*. Jadi, seluruh jaringan perlu dipertimbangkan untuk menentukan skor.

2. 4 Kajian Pustaka

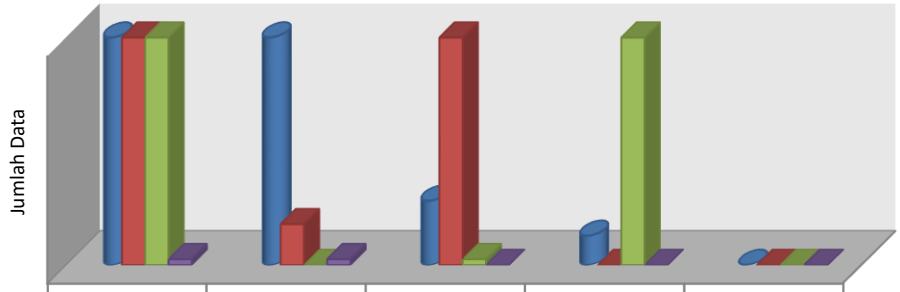
Terdapat berbagai teknik untuk menangani ketidakpastian dalam masalah klasifikasi yang telah dikembangkan seperti metode *Neural Network*. Akan tetapi, *Neural Network* sering terkendala pada masalah *overfitting*[8]. Selain itu juga ada teori Dempster-Shafer[9], namun bagaimanapun, kedua algoritma sebelumnya menggambarkan konsep hubungan secara *black-box* dan sulit untuk ditelusuri.

Dalam kasus diagnosis penyakit, beberapa algoritma juga telah digunakan sebelumnya dan salah satu yang cukup dikenal secara umum adalah algoritma *naïve bayes*[10,11]. Algoritma *naïve bayes* diklaim bekerja cukup cepat dan mengindentifikasi hubungan dengan probabilitas kejadian fitur pada masing-masing kelas, namun tidak mampu menggambarkan lebih jauh hubungan antar fitur-fitur dalam dataset, mengingat konsep hubungan antar fiturnya sejak awal adalah independen. Padahal, hubungan tersebut sangat penting untuk diagnosa dan meningkatkan keyakinan akan presisi pada prediksi. Penelitian ini memfasilitasi hal tersebut dengan metode *Bayesian Network* menggunakan konsep DAGnya. Sejauh ini, untuk mengoptimalkan pencarian, beberapa model telah diusulkan dalam pembelajaran struktur *Bayesian Network* dengan menggunakan algoritma heuristik dan metaheuristik. Salah satu studi terkait yang juga cukup relevan dengan penelitian ini adalah makalah[12]. Makalah tersebut mengevaluasi perilaku algoritma K2 untuk pembelajaran struktur *Bayesian Network* dari dataset terbatas yaitu data cuaca. Kesimpulan dari makalah tersebut bahwa ada penggunaan terbatas *Bayesian Network* dalam domain set data kecil.

3. HASIL DAN PEMBAHASAN

Di bagian ini, lima (5) algoritma dijalankan, yaitu algoritma BN-TAN sebagai metode yang diusulkan untuk menklasifikasi penyakit hipotiroid dan empat algoritma yang dijadikan sebagai pembanding. Hasil perbandingan klasifikasi setiap *instance* dalam 4 kelas yang ada pada dataset dapat dilihat pada Gambar 2. Berdasarkan gambar tersebut jumlah data yang terklasifikasi benar pada kelas negative adalah 3462 dari 3480 data, pada kelas compensated-hipotiroid

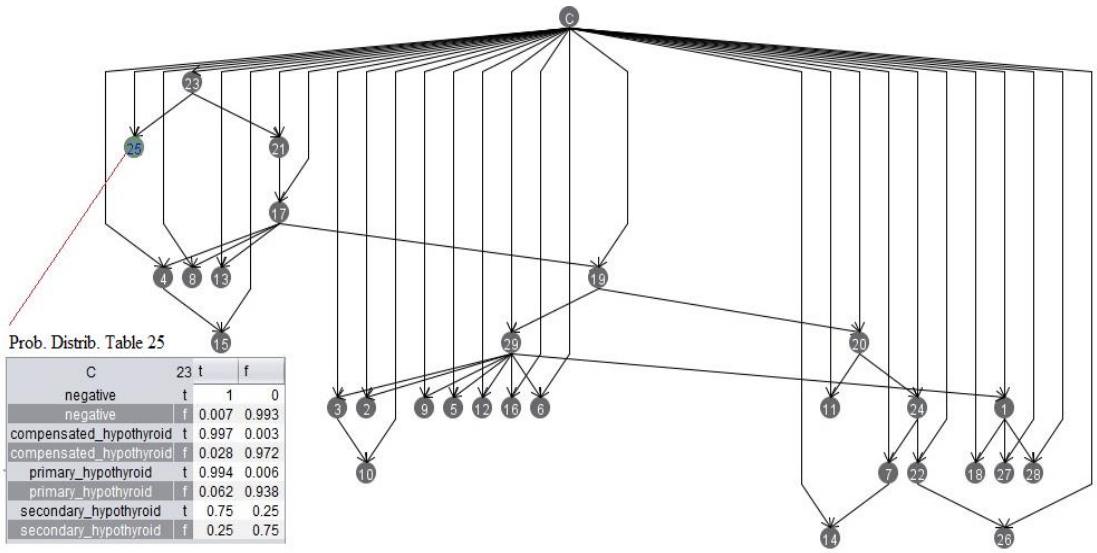
sebanyak 186 dari 194 data, pada kelas primary-hipotiroid terklasifikasi benar 92 dari 95 data, sementara kelas secondary-hipotiroid terjadi kesalahan klasifikasi untuk total 2 data. Ini terjadi karena kelompok data yang terbagi tidak seimbang, sehingga kecendrungan kelompok kecil mengikut pada kelas besar atau mayoritas. Namun secara keseluruhan akurasi yang dihasilkan BN-TAN adalah 93,13%.



Kelas Data	a=negative	b = compens	c = primary	d = secondary
a	3480	3462	12	6	0
b	194	8	186	0	0
c	95	1	2	92	0
d	2	2	0	0	0

Gambar 2 Grafik Confusion Matriks

Struktur BN-TAN pada dataset hipotiroid dipresentasikan dalam Gambar 3. Mengingat jumlah atribut adalah 30 dan nama atribut yang panjang maka pada graf struktur pembelajaran yang ditampilkan, nama disingkat menjadi bilangan 0 sampai 29. Dimana atribut ke 29 adalah atribut kelas (*class*). Beberapa hubungan antar atribut dapat ditampilkan dan selanjutnya dapat digunakan untuk menganalisis pola hubungan antar atribut pada penyakit hipotiroid. Setelah struktur jaringan yang baik diidentifikasi, tabel probabilitas bersyarat untuk masing-masing variabel dapat diperkirakan.



C	23	t	f
negative	t	1	0
negative	f	0.007	0.993
compensated_hypothyroid	t	0.997	0.003
compensated_hypothyroid	f	0.028	0.972
primary_hypothyroid	t	0.994	0.006
primary_hypothyroid	f	0.062	0.938
secondary_hypothyroid	t	0.75	0.25
secondary_hypothyroid	f	0.25	0.75

Gambar 3 Struktur BN dengan Tree Augmented Network pada dataset Hipotiroid

Selanjutnya dihitung hasil akurasi dan waktu konstruksi pada proses pengujian dengan membandingkan 5 algoritma yaitu Naïve Bayes, BN-TAN, BN-Hill Climbing (BN-HC), BN-K2 dan BN-Tabu Search (BN-TS).

Tabel 1: Kinerja klasifikasi (akurasi) dari lima algoritma yang diukur pada dataset hipotiroid.

Struktur Pembelajaran	***Akurasi Klasifikasi***	***Waktu Konstruksi Model***
Naïve Bayes	92.79 %	**0.01**
BN-TAN	**93.13 %**	0.06
BN-HC	92.89 %	0.2
BN-K2	92.63 %	0.04
BN-TS	92.84 %	0.1

BN-TAN memiliki nilai akurasi yang mampu bersaing dengan algoritma lain yang sejenis dan memiliki waktu konstruksi sebesar 0.06 detik, sebagaimana yang dapat dilihat pada Table 1. Kami mencatat akurasi klasifikasi dan waktu konstruksi untuk setiap algoritma secara detail dalam tabel yang sama. Hasil performa klasifikasi dari 5 (lima) algoritma tersebut, dapat dilihat bahwa Algoritma BN-TAN paling berbeda dari 4 (empat) algoritma lainnya. Untuk waktu konstruksi, meskipun BN-TAN bukan yang meraih waktu konstruksi terkecil, namun waktu yang digunakan berada diurutan ke 3 (tiga) terkecil dari 5 (lima) algoritma yang diperbandingkan.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari makalah ini adalah bahwa struktur pembelajaran *Bayesian Network* dengan algoritma BN-TAN mampu menggambarkan hubungan antar atribut dalam dataset hipotiroid. Selain dari pada itu, hasil validasi terhadap algorithma pencarian TAN yang digunakan pada struktur pembelajaran *Bayesian Network* menunjukkan bahwa BN-TAN meraih akurasi terbaik setelah membandingkannya dengan 4 algoritma kompetitif yang telah dijelaskan sebelumnya. Sementara untuk waktu konstruksi yang dibutuhkan BN-TAN menggunakan waktu yang sedang.

5. SARAN

Saran-saran penelitian lebih lanjut untuk penelitian ini adalah bahwa penelitian lanjutan dapat menyertakan analisis tambahan dengan mempertimbangkan masalah ketidakseimbangan kelas data pada dataset dan teknik diskritisasi yang sesuai. Lebih jauh, struktur pembelajaran dapat dibandingkan dengan algoritma lainnya.

DAFTAR PUSTAKA

[1] K. P. Livin, 2015, *Manifestasi Kelainan Tiroid Di Bagian Penyakit Dalam Rs. Dr. M. Djamil, Padang Pada Tahun 2014*, Upt. Perpustakaan Unand.

[2] H. C. Ong, 2011, Improving Classification In Bayesian Networks Using Structural Learning, *World Acad. Sci. Eng. Technol*, Vol. 75, Hal. 1407–1411.

[3] M. Lichman, 2013, *Uci Machine Learning Repository*, Irvine, Ca.

[4] R. R. Bouckaert, Bayesian Network Classifiers In Weka For Version 3-5-7, 2008, *Artif. Intell. Tools*, Vol. 11, No. 3, Hal. 369–387.

[5] I. H. Witten, E. Frank, M. A. Hall, Dan C. J. Pal, 2016, *Data Mining: Practical Machine Learning Tools And Techniques*. Morgan Kaufmann, 2016.

[6] A. S. Hesar, H. Tabatabaee, Dan M. Jalali, 2012, Structure Learning Of Bayesian Networks Using Heuristic Methods, *In Proc. Of International Conference On Information And Knowledge Management (Icikm 2012)*.

[7] P. Refaeilzadeh, L. Tang, Dan H. Liu, 2009, Cross-Validation, *Encycl. Database Syst*, Hal. 532–538.

[8] G. Panchal, A. Ganatra, P. Shah, Dan D. Panchal, 2011, Determination Of Over-Learning And Over-Fitting Problem In Back Propagation Neural Network, *Int. J. Soft Comput*, Vol. 2, No. 2, Hal. 40–51.

[9] R. R. Yager Dan N. Alajlan, 2015, Dempster–Shafer Belief Structures For Decision Making Under Uncertainty, *Knowledge-Based Syst*, Vol. 80, Hal. 58–66.

[10] Sabransyah, M. Dan Nasution, Yuki Novia; Amijaya, Fidia Deny Tisna, 2017, Aplikasi Metode Naive Bayes Dalam Prediksi Risiko Penyakit Jantung, *Jurnal Eksponensial*, Vol. 8 No.2, Hal. 11-118.

[11] Agustiawan, Beni, 2015 Sistem Klasifikasi Penyakit Tenggorokan Berbasis Web Menggunakan Metode Naive Bayes. *Skripsi*, Semarang: Universitas Dian Nuswantoro.

[12] H. Maharani, 2015, Konstruksi Bayesian Network Dengan Algoritma K2 Pada Kasus Prediksi Cuaca, *J. Telemat*, Vol. 7, No. 2.