

Penerapan Algoritma *Tree Augmented Naive Bayesian* pada Penentuan Peubah Penting

PINGKAN AWALIA¹, AJI HAMIM WIGENA², ANANG KURNIA³

¹Student of Statistics Department, Bogor Agricultural University, ²Lecture of Statistics Department, Bogor Agricultural University, ³ Lecture of Statistics Department, Bogor Agricultural University

ABSTRACT

In the era of free market competition today, improving product quality is very important. Consumer preferences through product level of analysis is one method that many manufacturers conducted to evaluate the product. Multivariable regression is a statistical method used to determine the important variables. The weakness of this method is the strict assumption. This problem will be completed by the method of bayesian networks. There are several algorithms to build the BN. This study uses TAN and NB because of its simplicity. This study shows that the most accurate method at the chosen level of classification accuracy is the TAN by 83%. The importance variable is the aspect liking of strength of after taste.

Keywords : Bayesian Network, Naive Bayesian, Tree Augmented Naive Bayesian

1. PENDAHULUAN

Perbaikan kualitas sangat penting dilakukan oleh setiap produsen secara kontinu agar konsumen tidak beralih ke produk lain. Peningkatan kualitas akan efektif jika produsen mengetahui prioritas atribut produk yang disukai oleh konsumen yang lebih dikenal dengan tingkat kepentingan peubah.

Banyak metode yang dapat digunakan untuk menentukan tingkat kepentingan peubah. Metode statistika yang selama ini sering digunakan oleh perusahaan riset pemasaran adalah analisis regresi berganda. Namun demikian, salah satu masalah dalam penerapan analisis regresi linier berganda adalah asumsi yang mengikat. Sejumlah asumsi harus dipenuhi ketika menggunakan analisis regresi berganda seperti bentuk hubungannya linier, sisaan adalah peubah acak yang bebas terhadap nilai peubah penjelas, sisaan menyebar normal dengan rata-rata nol dan ragam yang konstan, homogen, dan antar sisaan tidak saling berkorelasi. Hal ini menyebabkan analisis regresi berganda menjadi kurang layak diterapkan karena asumsi tersebut sulit dipenuhi terlebih dalam banyak aplikasinya peubah respon diukur dalam skala ordinal. Salah satu alternatif yang dapat dilakukan untuk mengatasi kendala tersebut menggunakan metode *Bayesian network* (BN)..

BN merupakan salah satu metode model peluang grafis yang digunakan untuk klasifikasi, pemeringkatan, dan prediksi (Friedman *et al.* 1997). Metode ini cocok diterapkan pada data diskrit. Manfaat metode BN pada bidang statistika adalah mampu menggambarkan hubungan ketergantungan suatu peubah dari peubah yang lain sehingga dapat digunakan untuk mempelajari hubungan sebab akibat antar peubah.

Metode BN selain diterapkan pada bidang pemasaran juga diterapkan di *data mining*, bidang kesehatan, dan meteorologi. Aplikasi dalam bidang kesehatan digunakan untuk menentukan peluang seseorang mengalami suatu penyakit dari sebab-sebab yang mungkin. Bidang meteorologi digunakan untuk menghasilkan prakiraan peluang kejadian hujan di suatu stasiun. Bidang demografi untuk penentuan daerah tertinggal serta untuk memprediksi pemenang pilkada.

Ada beberapa algoritma untuk menyusun BN. Algoritma yang paling sederhana adalah *naive Bayesian* (NB). Metode ini mengasumsikan antar peubah penjelas tidak ada korelasi. Selain algoritma NB algoritma yang cocok diterapkan adalah *tree augmented naive Bayesian* (TAN) yang merupakan pengembangan dari metode NB. Algoritma ini bisa mengakomodir korelasi antar peubah penjelas (Friedman *et al.* 1997). Tujuan dari penelitian ini adalah: (1)

Menerapkan algoritma TAN pada BN dalam penentuan tingkat kepentingan peubah penjelas, (2) Membandingkan algoritma TAN dan NB, dan (3) Menganalisis perubahan peubah penjelas akibat perubahan peubah respon.

2. TINJAUAN PUSTAKA

2.1 Tingkat Kepentingan Peubah

Tingkat kepentingan peubah penjelas berguna untuk mengindikasikan tingkat kepentingan relatif dari masing-masing peubah dalam menduga model. Nilai yang digunakan merupakan nilai relatif sehingga hasil penjumlahannya adalah satu. Tingkat kepentingan peubah tidak berhubungan dengan akurasi model. Tingkat kepentingan berhubungan dengan kepentingan peubah penjelas dalam memprediksi model (Friedman *et al.* 1997).

Perhitungan tingkat kepentingan peubah lebih lama daripada pembentukan model, terlebih jika menggunakan data dengan jumlah yang besar. Tingkat kepentingan peubah berhubungan dengan peubah yang mereduksi keragaman dari peubah respon.

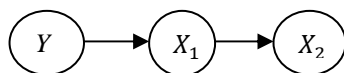
2.2 Bayesian Network

Bayesian network (BN) adalah suatu metode yang didasarkan atas teorema Bayes. Teorema Bayes menyatakan jika S suatu ruang contoh dan $\{X_1, \dots, X_n\}$ merupakan sekatan S dengan syarat $P(X_i) \neq 0$, dan $i=1, \dots, n$. Jika Y merupakan suatu kejadian pada ruang contoh S dengan syarat $P(Y) \neq 0$, maka secara matematis kaidah peluang Bayes dapat dituliskan sebagai berikut :

$$P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{\sum_{i=1}^n P(X_i)P(Y|X_i)} \quad (1)$$

(Nasoetion 1984).

BN dapat memberikan informasi yang sederhana dan padat mengenai informasi peluang. BN berdasarkan komponennya terdiri dari struktur Bayesian dan parameter Bayesian. Struktur Bayesian merupakan sebuah *Directed Acyclic Graph* (DAG) yang menggambarkan hubungan ketergantungan antar peubah. Parameter Bayesian merupakan himpunan dari parameter setiap peubah berdasarkan graf tersebut. Struktur Bayesian terdiri dari simpul yang merepresentasikan peubah-peubah dan sisi yang merepresentasikan hubungan ketergantungan antar simpul. Setiap simpul yang dihubungkan menunjukkan hubungan ketergantungan. Misalkan himpunan dari simpul dinyatakan dengan $\{X_1, \dots, X_n\}$. Jika terdapat sisi dari simpul X_j ke simpul X_k , dikatakan bahwa X_j adalah *parent* dari X_k dan X_k adalah *child* bagi X_j . Himpunan *parent* dari simpul X_i dinotasikan sebagai Π_i . Contoh dapat dilihat pada Gambar 1 *parent* untuk X_2 adalah X_1 dan *child* untuk Y adalah X_2 dan X_1 .



Gambar 1 *Directed Acyclic Graph*.

Hasan (2007) menyatakan bahwa BN merupakan pasangan dari (G,P) dengan $G=(N,E)$ adalah DAG atau graf berarah yang tidak memiliki siklus dengan simpul (N) sebagai peubah dan sisi (E) sebagai penghubung kebebasan bersyarat antar peubah, sedangkan P adalah sebaran peluang bersyarat yang disebut parameter Bayesian. *Parents* didefinisikan sebagai simpul yang dijadikan syarat dan *child* adalah simpul yang diberikan syarat.

BN merupakan suatu metode berdasarkan teorema Bayes yang menggambarkan hubungan bersyarat :

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (2)$$

dengan $P(X|Y)$ disebut peluang posterior adalah peluang X setelah Y terjadi. $P(Y|X)$ disebut *likelihood* adalah peluang Y terjadi setelah X terjadi. $P(X)$ disebut juga prior adalah peluang kejadian X. $P(Y)$ adalah peluang kejadian Y dan $P(Y) \neq 0$.

Ada beberapa algoritma penyusun BN yaitu *naive Bayesian* dan *semi naive Bayesian*. Penelitian kali ini menggunakan algoritma TAN dan NB karena alasan keefektifan dan kesederhanaannya.

Pembuatan model dalam BN melibatkan dua langkah yaitu:

1. Membuat struktur jaringan.
Struktur jaringan dalam BN dibentuk dalam suatu graf. Sebuah graf terdiri dari kumpulan simpul dan sisi. Sisi merupakan garis yang menghubungkan titik simpul tersebut.
2. Menduga nilai peluang setiap simpul.

2.3 Semi Naive Bayesian

Metode *semi naive Bayesian* secara garis besar terdiri dari dua kelompok. Kelompok pertama membangun NB menggunakan sekumpulan peubah baru yang dihasilkan dari proses *deleting attributes* dan *joining attributes*. Kelompok yang kedua membangun metode *semi naive Bayesian* dari struktur garis penghubung secara jelas di antara peubah-peubah penjelas yang menunjukkan hubungan ketidakbebasan (saling mempengaruhi) antar peubah penjelas (Zheng & Webb 2005).

2.4 Tree Augmented Naive Bayesian

NB merupakan salah satu algoritma pembangun struktur BN dengan mengasumsikan antar peubah penjelas saling bebas. Cara kerja metode ini sangat sederhana yaitu menghubungkan semua peubah penjelas ke peubah respon. Arah sisi pada BN semua mengarah ke peubah respon karena asumsi *naive* menyatakan bahwa peubah penjelas diasumsikan saling bebas. Struktur BN yang dibangun dengan algoritma ini tidak ada hubungan antar peubah penjelas satu dengan penjelas yang lain.

TAN dikembangkan sebagai modifikasi dari metode Bayesian sederhana. Hal ini memungkinkan tiap peubah penjelas terhubung dengan peubah penjelas yang lain. Hal ini dapat meningkatkan keakuratan prediksinya. TAN (Friedman *et al.* 1997) menunjukkan kinerja yang baik meskipun sederhana. Misal $\Omega = \{X_1, \dots, X_m\}$ sebagai domain diskrit, X_1, \dots, X_m sebagai atribut dalam domain Ω . DAG menggambarkan hubungan bebas bersyarat. Tiap peubah X_i bebas dari bukan keturunannya berdasarkan dari *parents* yang digunakan.

TAN disusun menggunakan *Weighted Maximum Spanning Tree* (WMST). Metode ini menghubungkan masing-masing sisi dengan memanfaatkan informasi antar peubah. Saat matriks pembobotnya dibuat, algoritma MWST menghubungkan pohon yang dihubungkan dengan pilihan akar yang digunakan.

Penentuan akar sangat penting untuk meningkatkan keakuratan klasifikasi. Penentuan akar didasarkan pada peubah yang memiliki informasi bersama yang maksimum dirumuskan (Friedman *et al.* 1997)

$$X_{root} = \arg \max I_p(X_i, Y)$$

dengan X_{root} adalah peubah yang menjadi akar, $I_p(X_i, Y)$ adalah *mutual information*.

Langkah kerja algoritma TAN adalah sebagai berikut (Chow & Liu, 1968):

1. Menentukan data, peubah respon, dan peubah penjelas. Seluruh peubah dalam keadaan diskrit.
2. Membuat struktur pohon menggunakan algoritma penyusun strukturnya.
Langkah pembentukan struktur TAN sebagai
 - a. Hitung $I(X_i, X_j | Y), i = 1, \dots, n, j = 1, \dots, n, i \neq j$ antar masing-masing pasangan peubah.
 - b. Gunakan algoritma Prim untuk menyusun WMST dengan pembobotnya dari hubungan antar sisi nya X_i dengan X_j melalui $I(X_i, X_j | Y)$.
Langkah-langkah penyusunan algoritma Prim (Prim 1957)
 - b.1. Memilih peubah yang dijadikan sebagai input, dalam hal ini yang dipilih yaitu peubah respon.
 - b.2. Mencari peubah dengan bobot maksimal lalu hubungkan. Bobot maksimal menggunakan informasi bersama bersyarat (*conditional mutual information*).
 - b.3. Proses ini diulang sampai semua peubah dihubungkan.
 - c. Menambahkan peubah respon sebagai *parent* untuk masing-masing X_i dengan $1 \leq i \leq n$, n adalah jumlah peubah penjelas yang digunakan.
3. Mentransformasi hasil pohon yang belum berarah menjadi pohon berarah dengan memilih X_1 sebagai *root* dan panah mengarah keluar.
4. Menghitung peluang bersyarat masing-masing peubah.

2.5 Weighted Maximum Spanning Tree

Weighted Maximum Spanning Tree (WMST) merupakan bagian dari graf G yang tidak memiliki siklus namun memiliki simpul yang sama seperti G (Munir 2003). Syarat dalam membangun WMST adalah semua simpul terhubung dan memiliki bobot. Pembobot yang digunakan adalah *conditional mutual information*. Ada dua jenis algoritma dalam menyusun WMST yaitu algoritma Prim dan Kruskal. Penelitian kali ini menggunakan algoritma Prim.

Mutual information (Friedman *et al.* 1997) disebut juga ukuran kedekatan dalam mengaproksimasi X_i dengan X_j yang merupakan dasar dari $I(X_i, X_j)$ adalah perbedaan antara informasi yang terkandung dalam X_i dan X_j . Informasi bersama antara dua node X_i dan X_j dirumuskan

$$I(X_i, X_j) = \sum_{x_i, x_j} \Pr(x_i, x_j) \log\left(\frac{\Pr(x_i, x_j)}{\Pr(x_i) \Pr(x_j)}\right) \quad (3)$$

Informasi bersama yang dibentuk digantikan oleh *conditional mutual information* antara dua peubah penjelas dan target yang digunakan. Rumus yang digunakan adalah

$$I(X_i, X_j | Y) = \sum_{x_i, x_j, y_k} \Pr(x_i, x_j, y_k) \log\left(\frac{\Pr(x_i, x_j | y_k)}{\Pr(x_i | y_k) \Pr(x_j | y_k)}\right) \quad (4)$$

rumus ini menunjukkan informasi bersama antara X_i dan X_j jika Y diketahui.

2.6 Evaluasi Klasifikasi Bayesian Network

Klasifikasi merupakan penerapan kaidah Bayes dengan memprediksi peluang posterior terbesar. Evaluasi klasifikasi BN dilakukan dengan membandingkan hasil nilai prediksi struktur dengan nilai aktualnya. Struktur tersebut memperlihatkan tingkat yang digunakan. Jika nilai $P(X_k | Y)$ untuk semua $k=1,2,3,\dots,n$ diperoleh maka peluang masuk kategori ke- k adalah peluang terbesar dari dugaan klasifikasi. Rumus yang digunakan adalah

$$P(X_k | Y) = \arg \max P(Y | X_i) P(X_i) \quad (5)$$

Evaluasi klasifikasi struktur BN dapat dilakukan dengan membandingkan nilai hasil prediksi dengan nilai aktual. Evaluasi ini menghasilkan tingkat keakuratan (Purwadi 2009) dengan rumus sebagai berikut

$$\text{Tingkat keakuratan} = \frac{\text{prediksi sesuai}}{\text{prediksi total}} \quad (6)$$

3. METODOLOGI

3.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapatkan dari perusahaan riset pemasaran tahun 2011. Data ini merupakan data kesediaan responden membeli suatu produk makanan ringan. Data terdiri dari satu produk makanan dengan jumlah responden sebanyak 200 yang merupakan anak-anak dengan kriteria umur 8-13 tahun. Data yang digunakan terdiri dari delapan peubah. Masing-masing peubah terdiri dari sembilan kategori. Keterangan kategori peubah respon dan peubah penjelas dapat dilihat pada Tabel 1.

Tabel 1 Kategori peubah penjelas dan peubah respon

Kode	Keterangan
1	Sama sekali tidak suka
2	Sangat tidak suka
3	Lumayan tidak suka
4	Agak tidak suka
5	Biasa saja
6	Agak suka
7	Lumayan suka
8	Sangat suka
9	Sangat suka sekali

Peubah yang digunakan adalah:

1. Keseluruhan kesukaan (Y)
2. Warna produk (X_1)
3. Kekuatan aroma produk (X_2)
4. Ketebalan produk (X_3)
5. Kelembutan produk (X_4)
6. Rasa Asin pada produk (X_5)
7. Kekuatan rasa susu produk (X_6)
8. Kelezatan produk (X_7)
9. Kekuatan rasa produk setelah dirasakan (X_8).

3.2 Metode

Metode penelitian dilakukan dengan cara membangun struktur BN menggunakan dua buah algoritma. Algoritma yang digunakan adalah NB dan TAN. Kedua struktur BN akan dihitung tingkat keakuratan klasifikasi. Tahap-tahap yang dilakukan adalah :

1. Menentukan peubah yang akan digunakan, jika peubah berupa data kontinu perlu didiskritkan terlebih dahulu.
2. Menyusun struktur BN menggunakan algoritma TAN untuk menentukan ukuran tingkat kepentingan dan algoritma NB.
3. Menentukan parameter BN menggunakan algoritma TAN dan NB.
4. Menghitung nilai keakuratan klasifikasi metode TAN dan NB.
5. Menyusun dan menentukan parameter untuk tiap-tiap akar yang dipilih.
6. Menganalisis perubahan peluang peubah respon berdasarkan perubahan peluang yang terjadi pada setiap peubah penjelas dan sebaliknya berdasarkan struktur BN yang terbaik.

4. HASIL DAN PEMBAHASAN

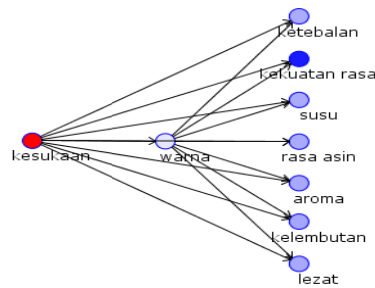
4.1 Algoritma TAN

Algoritma TAN merupakan pengembangan dari algoritma NB. Algoritma NB mengasumsikan antar peubah penjelas tidak berkorelasi. Pada penelitian kali ini terdapat korelasi antar peubah penjelas. Hal ini terlihat dari nilai korelasi Spearman hampir semua peubah memiliki hubungan yang erat.

Struktur algoritma TAN dibangun menggunakan *weighted maximum spanning tree*. WMST dalam membangun pohonnya menggunakan algoritma Prim. Algoritma yang digunakan mensyaratkan akar dan peubah respon panahnya mengarah keluar. Pemilihan akar yang digunakan tertera di metode. Penentuan hubungan antar peubah penjelas dengan cara memaksimalkan *conditional mutual information* antar peubah penjelas bersyarat peubah respon.

Gambar 2 memperlihatkan struktur BN yang dibangun menggunakan algoritma TAN. Algoritma ini mensyaratkan hubungan tidak membentuk siklus. Peubah respon yang digunakan adalah keseluruhan kesukaan, peubah yang lain sebagai peubah penjelas. Peubah yang menjadi *parent* untuk semua peubah adalah peubah respon. Peubah yang menjadi akar (*root*) adalah peubah X_1 yaitu warna produk. Jika kita mengubah akarnya maka tingkat akurasi akan berubah.

Rumus perhitungan manual untuk penentuan tingkat kepentingan peubah belum tersedia. Hal ini karena perhitungan rumusnya membutuhkan waktu yang lebih lama daripada menyusun model. Gambar 3 memperlihatkan tentang ukuran tingkat kepentingan dari peubah penjelas terhadap peubah respon. Gambar 3 memperlihatkan bahwa peubah yang paling mempengaruhi keseluruhan kesukaan dan yang harus ditingkatkan adalah kekuatan rasa produk setelah dirasakan.



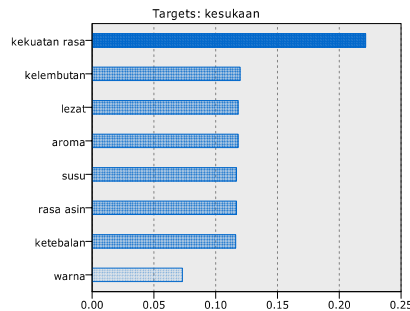
Gambar 2 Struktur BN menggunakan akar warna produk.

Peringkat kedua yang perlu ditingkatkan adalah tingkat kelembutan produk. Urutan ketiga adalah tingkat kelezatan produk. Produk yang dicobakan mungkin sudah lezat namun perlu ditingkatkan lagi komposisi bahan-bahan penambah kelezatan. Urutan keempat adalah tingkat kekuatan aroma produk. Urutan kelima adalah kekuatan rasa susu. Ada sebagian besar responden yang suka bila kandungan susu ditingkatkan. Ada pula yang tidak suka dengan kandungan susu pada produk yang dicobakan.

Urutan keenam adalah tingkat keasinan produk. Ketujuh adalah tingkat ketebalan produknya. Tingkat kepentingan paling rendah terhadap peubah respon adalah warna dari produk. Warna produk yang ditawarkan ternyata tidak dipentingkan oleh responden. Hal ini dikarenakan responden sudah puas dengan tingkat warna pada produk. Tingkat kepentingan ini menunjukkan semakin ke bawah tingkat kepentingannya semakin kecil.

Tabel 2 memperlihatkan dugaan klasifikasi menggunakan algoritma TAN dengan pemilihan akar adalah warna produk. Tingkat keakuratan yang dicapai sebesar 79%.

Penentuan akar dari perangkat lunak yang digunakan adalah X_1 . Tahap selanjutnya mengganti akarnya menggunakan peubah penjelas yang lain untuk meningkatkan tingkat akurasinya.



Gambar 3 Ukuran tingkat kepentingan peubah penjelas menggunakan akar warna produk.

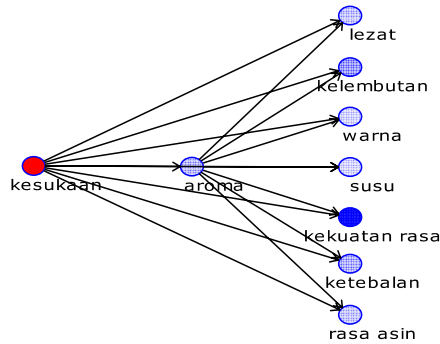
Tabel 2 Dugaan klasifikasi menggunakan akar warna produk

		Aktual					Total
		4	6	7	8	9	
Dugaan	4	1	0	0	0	0	1
	6	0	2	0	0	0	2
	7	0	0	3	3	1	7
	8	0	1	12	134	14	161
	9	0	0	1	10	18	29
Total		1	3	16	147	33	200

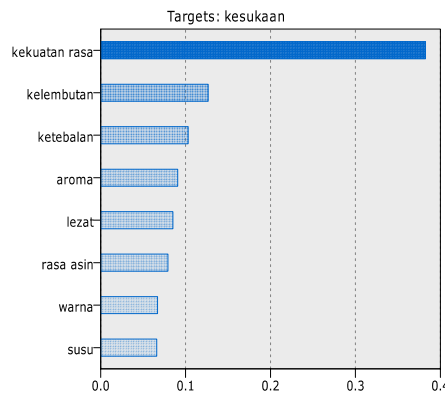
Tingkat akurasi struktur BN mencapai nilai tertinggi saat akar yang digunakan adalah kekuatan aroma produk. Gambar 4 dari struktur BN menggunakan akar kekuatan aroma produk.

Gambar 5 memperlihatkan bahwa peubah yang penting atau yang harus ditingkatkan sama dengan struktur BN dengan akar warna produk. Tingkat kepentingan peubah yang paling kecil

adalah kekuatan rasa susu. Tingkat kepentingan relatif sama menggunakan akar warna produk.



Gambar 4 Struktur BN menggunakan akar kekuatan aroma produk.



Gambar 5 Ukuran tingkat kepentingan peubah penjelas menggunakan akar kekuatan aroma produk.

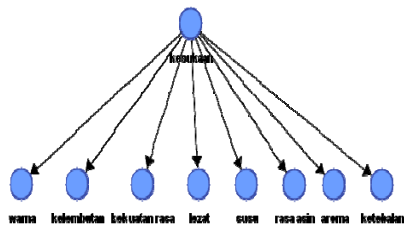
Tabel 3 memperlihatkan dugaan klasifikasi menggunakan algoritma TAN dengan pemilihan akar adalah kekuatan aroma produk. Pemilihan akar berdasarkan tingkat akurasi yang dihasilkan. Tingkat keakuratan yang dicapai sebesar 83%.

Tabel 3 Dugaan klasifikasi menggunakan akar kekuatan aroma produk

		Aktual					Total
		4	6	7	8	9	
Dugaan	4	1	0	0	0	0	1
	6	0	3	1	1	0	5
	7	0	0	10	6	0	16
	8	0	0	5	130	11	146
	9	0	0	0	10	22	23
Total		1	3	16	147	33	200

4.2 Algoritma NB

Algoritma NB merupakan salah satu metode pengklasifikasian. Metode ini berdasarkan penerapan teorema Bayes dengan asumsi antar peubah saling bebas. Peluang prior yang digunakan adalah frekuensi masing-masing kategori untuk peubah respon. Arah panah target menuju ke peubah penjelas.

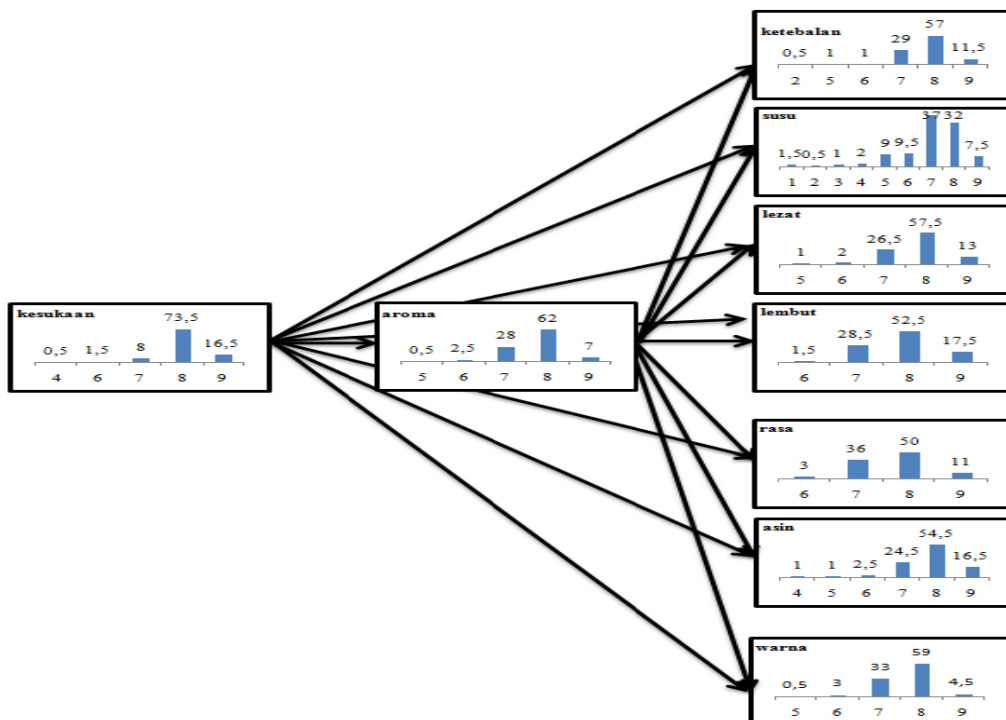


Gambar 6 Struktur BN menggunakan algoritma NB.

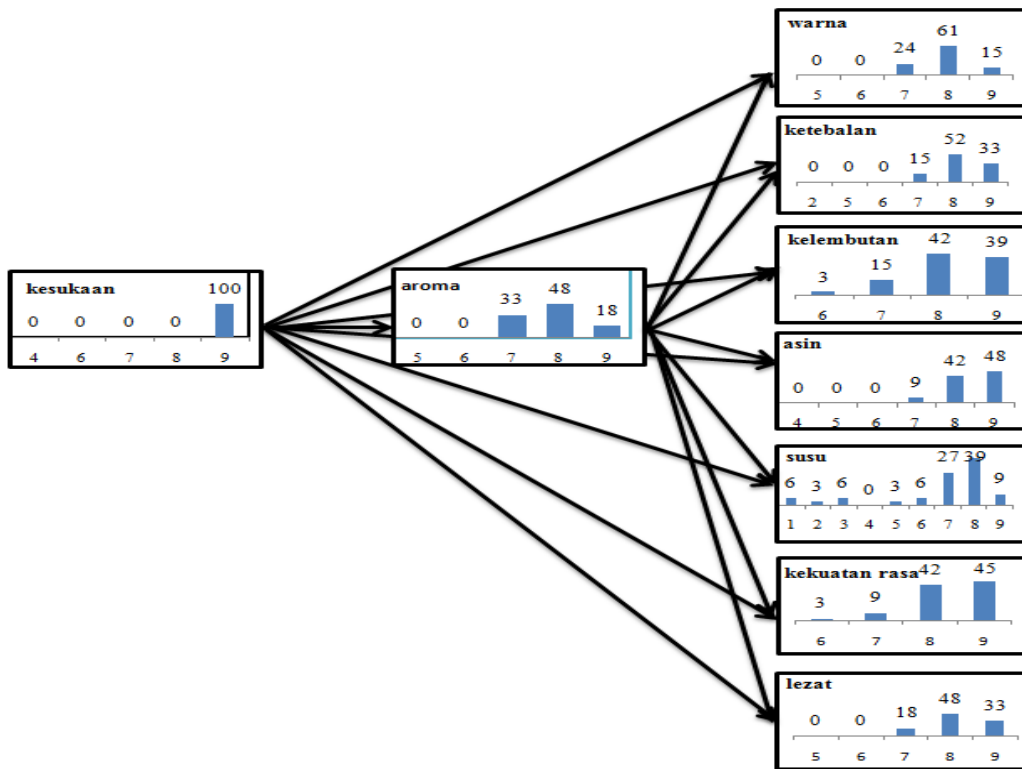
Tingkat keakuratan klasifikasi untuk algoritma NB sebesar 71.5% dengan dugaan dapat dilihat pada Tabel 4.

Tabel 4 Dugaan klasifikasi menggunakan algoritma NB

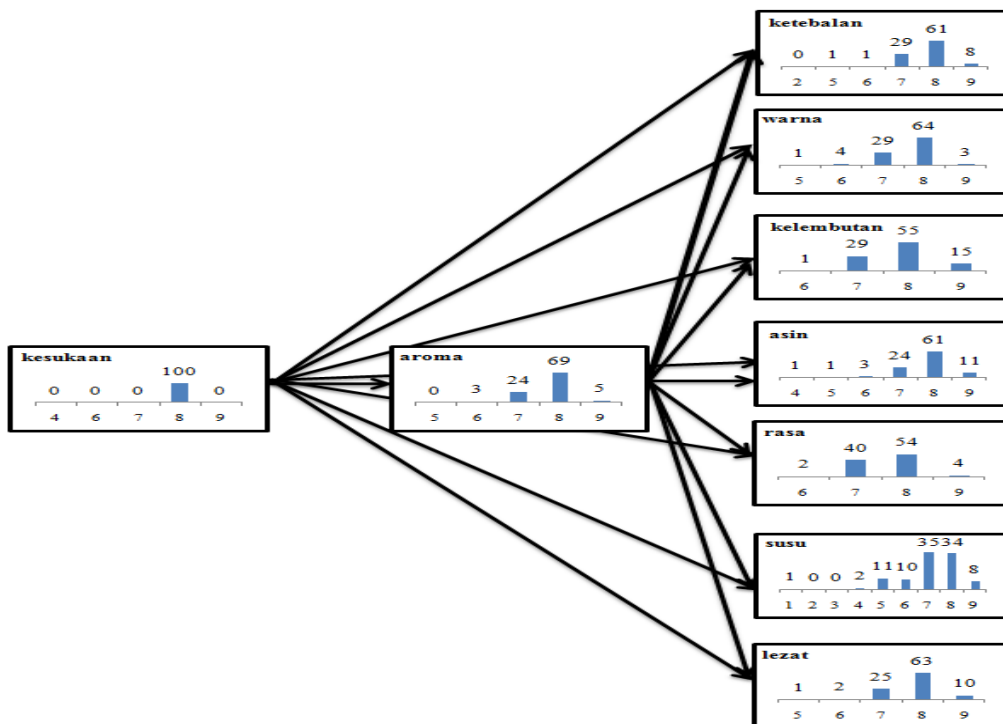
	Aktual					Total
	4	6	7	8	9	
Dugaan	4	1	0	0	0	1
	6	0	2	0	0	2
	7	0	0	8	17	27
	8	0	1	7	114	133
	9	0	0	1	16	35
Total	1	3	16	147	33	200



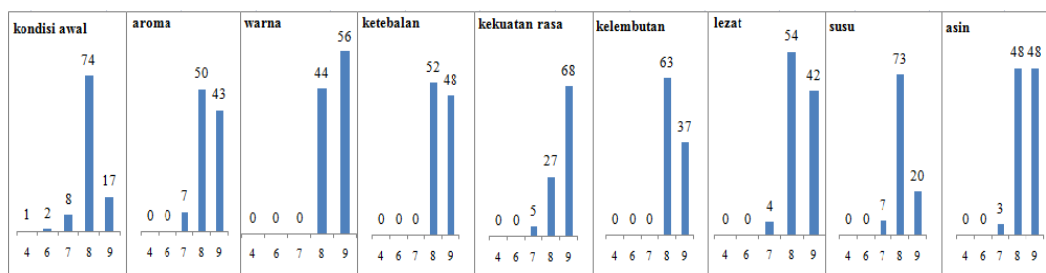
Gambar 7 Kondisi awal (%) peubah penjasar dan peubah respon.



Gambar 8 Kondisi (%) peubah-peubah penjelas saat kategori sangat suka sekali (9) pada tingkat kesukaan (Y) ditingkatkan menjadi 100%.



Gambar 9 Kondisi (%) peubah penjelas saat kategori sangat suka (8) pada tingkat kesukaan ditingkatkan menjadi 100%.



Gambar 10 Perubahan (%) yang terjadi pada peubah respon saat kategori sangat suka sekali (9) pada masing-masing peubah penjelas ditingkatkan menjadi 100%.

4.3 Penentuan Parameter BN

Peluang prior yang dipilih pada penelitian ini adalah frekuensi dari masing-masing kategori pada peubah respon. Parameter dalam BN berupa peluang bersyarat antara *parents* dan *child*. Contoh peluang bersyarat untuk peubah warna produk adalah kategori 5 dengan syarat kategori 4 peluang bersyaratnya 0, berturut-turut untuk kategori 6, 7, 8, dan 9 nilainya adalah 0, 1, 0, dan 0. Peluang bersyarat ini diperoleh menggunakan rumus yang terdapat pada Persamaan 1. Peluang bersyarat pada X_1 merupakan peluang X_1 dengan syarat Y sedangkan peubah yang lain yang menjadi *parents* adalah X_1 dan Y .

4.4 Analisis Perubahan Peubah Penjelas

Model dugaan klasifikasi BN dapat digunakan untuk melihat perubahan peluang yang terjadi pada peubah respon ketika peluang peubah penjelas berubah. Hal ini dapat dilakukan dengan cara mengubah peluang pada kategori-kategori peubah penjelas dan dilihat sejauh mana perubahan peluang yang terjadi pada peubah respon. Struktur BN yang digunakan adalah struktur yang memiliki tingkat akurasi yang tertinggi yaitu TAN dengan akar kekuatan aroma produk.

Pembahasan selanjutnya difokuskan pada analisis perubahan persentase pada kategori sangat suka (8) dan sangat suka sekali (9), baik pada peubah penjelas maupun peubah respon. Tabel 5 memperlihatkan contoh perubahan yang terjadi pada peubah tingkat kesukaan (Y) saat persentase kategori pada peubah kekuatan aroma produk (X_1), warna produk (X_2), dan semua peubah penjelas ditingkatkan menjadi 100%. Jika kategori sangat suka (8) pada peubah penjelas ditingkatkan menjadi 100% berakibat peningkatan persentase kategori sangat suka (8) dan persentase kategori sangat suka sekali (9) pada tingkat kesukaan (Y). Saat persentase kategori sangat suka sekali (9) pada peubah penjelas ditingkatkan menjadi 100% berakibat pada penurunan persentase kategori sangat suka (8) dan peningkatan persentase sangat suka sekali (9) pada peubah tingkat kesukaan (Y).

Perubahan persentase kategori-kategori peubah respon saat kategori sangat suka sekali (9) pada setiap peubah penjelas ditingkatkan menjadi 100% dapat dilihat pada Gambar 10. Besarnya kontribusi peubah penjelas terhadap perubahan persentase kategori sangat suka sekali (9) pada tingkat kesukaan (Y) disajikan pada Tabel 6. Nilai tersebut diperoleh dari selisih persentase kategori sangat suka sekali (9) pada peubah respon saat persentase kategori sangat suka sekali (9) dan sangat suka (8) peubah penjelas ditingkatkan menjadi 100%.

Tabel 6 memperlihatkan bahwa peubah penjelas yang memiliki kontribusi paling besar adalah peubah kekuatan rasa setelah dirasakan dengan kontribusi sebesar 22.95%.

Peubah rasa susu merupakan peubah yang memiliki kontribusi terkecil terhadap persentase kategori sangat suka sekali (9) yaitu sebesar -1.31%. Peubah rasa susu bernilai negatif, artinya saat kategori sangat suka (8) ditingkatkan menjadi kategori sangat suka sekali (9) kategori sangat suka sekali (9) pada peubah tingkat kesukaan (Y) mengalami penurunan.

Tabel 6 Kontribusi peubah penjelas

Peubah Penjelas	Persentase Kontribusi
Kekuatan rasa produk setelah dirasakan	22.95
Warna produk	16.35
Rasa asin produk	15.09
Ketebalan produk	13.94
Kelezatan produk	12.03
Kelembutan produk	10.08
Kekuatan aroma produk	9.67
Rasa susu produk	-1.31

4.5 Analisis Perubahan Peubah Respon

Analisis perubahan peubah penjelas terhadap respon sudah dijelaskan pada pembahasan sebelumnya. Adapun analisis perubahan peubah respon meneliti kondisi-kondisi peubah penjelas yang dapat menyebabkan peubah respon berada pada kondisi tertentu.

Kondisi ini memperlihatkan bahwa konsumen sangat menyukai produk yang ditawarkan oleh produsen. Gambar 8 memperlihatkan kondisi peubah-peubah penjelas saat kategori sangat suka sekali (9) ditingkatkan menjadi 100%.

Perubahan persentase terbesar setelah kategori sangat suka sekali (9) ditingkatkan menjadi 100% terjadi pada peubah rasa asin pada produk (X₆), yaitu sebesar 48.48%. Hal ini menunjukkan bahwa peubah penjelas tersebut cukup sensitif terhadap perubahan pada peubah respon.

Peubah yang paling kecil pengaruhnya jika kategori sangat suka sekali pada peubah respon ditingkatkan menjadi 100% adalah rasa susu. Persentase kondisi kategori sangat suka (8) pada tingkat kesukaan (Y) mencapai 100% peubah yang memiliki persentase kategori sangat suka (8) terbesar adalah aroma dari produk. Peubah yang paling kecil pengaruhnya adalah rasa susu.

Tabel 5 Perubahan persentase peubah respon saat persentase kategori sangat suka (8) dan sangat suka sekali (9) pada peubah warna produk, kekuatan aroma produk produk, dan semua peubah penjelas ditingkatkan menjadi 100%

	Tingkat Kesukaan (Y) (%)	
	Sangat suka (8)	Sangat suka sekali (9)
Persentase awal	73.5	16.5
Ditingkatkan menjadi 100%		
Warna Produk	Sangat suka (8) Sangat suka sekali (9)	79.66 44.44
Ditingkatkan menjadi 100%		
Kekuatan aroma produk	Sangat suka (8) Sangat suka sekali (9)	81.45 50
Ditingkatkan menjadi 100%		
Semua Peubah Penjelas	Sangat suka (8) Sangat suka sekali (9)	95.41 9.56

5. KESIMPULAN DAN SARAN

Tingkat keakuratan algoritma TAN lebih tinggi daripada algoritma NB sehingga algoritma TAN lebih tepat digunakan. Pemilihan akar yang memberikan tingkat keakuratan paling tinggi adalah kekuatan aroma produk dan rasa asin. Peubah yang paling penting untuk ditingkatkan adalah kekuatan rasa setelah dirasakan dan tingkat kepentingan paling rendah adalah rasa susu. Jika presentase kategori sangat suka sekali pada peubah penjelas ditingkatkan menjadi 100% maka presentase kategori sangat suka sekali pada peubah respon mengalami peningkatan tinggi.

Algoritma yang digunakan pada penelitian ini merupakan salah satu dari algoritma untuk membangun BN. Penelitian lebih lanjut diperlukan untuk membandingkan antara algoritma TAN dengan *SuperParent* TAN dan kategorisasi tiga skala.

DAFTAR PUSTAKA

- [1]. Chow CK, Liu CN. 1968. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory* 14: 462-467.
- [2]. Friedman N, Geiger D, Goldszmidt M. 1997. Bayesian network classifiers. *Machine Learning*, 29, 131-136.
- [3]. Hassan A. 2007. A probabilistic relaxation framework for learning bayesian network structures from data [tesis]. Kairo: Faculty Of Engineering, Cairo University.
- [4]. Neapolitan RE. 2004. *Learning Bayesian Network*. Northeast Illinois University Chicago : Pearson Prentice Hall.
- [6]. Purwadi I. 2009. Penerapan bayesian network dalam penetapan daerah tertinggal [skripsi]. Bogor: Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- [7]. Prim RC. 1957. Shortest connection networks and some generalisations. *Bell System Technical Journal* 36: 1389-1401.
- [8]. Zheng F, Webb GI. 2005. A comparative study of semi-naive bayes methods in classification learning. Di dalam Simoff SJ, Williams GJ, Galloway J, Kolyshkina I (Eds.). *Proceedings of Fourth Australasian Data Mining Conference (AusDM05)*; Sydney, Australia. Sydney: University of Techno. hlm 141-156.