

**MODEL PENGENALAN TERBAIK
DENGAN *TREE-AUGMENTED NETWORK* (TAN)
DAN *ESTIMATOR MAXIMUM LIKELIHOOD* (ML)
BERDASARKAN FITUR OBJEK**

Irwan Budi Santoso

Jurusan Teknik Informatika, Sains dan Teknologi
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
irwan.budi331177@gmail.com

Abstrak-Pengenalan suatu objek sangat tergantung dari seberapa handal model yang digunakan serta parameter model tersebut. *Tree-Augmented Network* (TAN) adalah salah satu model yang handal dalam melakukan klasifikasi, yang dibangun dengan memperhatikan hubungan diantara pasangan fitur-fitur objek. Sedangkan *Maximum Likelihood* (ML) adalah salah satu estimator yang telah banyak digunakan dan telah teruji penggunaannya. Kehandalan model serta estimator yang digunakan pada kenyataan belum cukup untuk menghasilkan model pengenalan terbaik, akan tetapi ada faktor lain yang memberi kontribusi besar yaitu dimensi atau fitur objek yang digunakan dalam membangun model tersebut. Hasil eksperimen untuk data training yang terdiri dari 5 jenis objek ringan menunjukkan untuk dimensi objek 5x5 (25 fitur), 6x6 (36 fitur) dan 7x6 (42 fitur) menghasilkan model TAN terbaik karena memberikan tingkat akurasi sistem 100%, sedangkan untuk dimensi dibawah atau diatas tersebut menghasilkan tingkat akurasi sistem yang lebih rendah. Kondisi tersebut menunjukkan bahwa tidak selamanya semakin besar dimensi atau fitur objek yang digunakan dalam training akan menghasilkan model yang semakin baik, karena bisa jadi semakin besar dimensi atau fitur objek akan menghasilkan informasi overflow.

Kata Kunci : *Tree-Augmented Network*, *Maximum Likelihood*, *Dimensi (fitur) objek*

1. PENDAHULUAN

Akurasi dalam mengenali suatu objek sangat tergantung dari seberapa handal model yang digunakan dalam pengenalan serta *estimator* yang digunakan untuk membangun dan mengestimasi parameter model tersebut. *Tree-Augmented Network* (TAN) adalah salah satu model yang handal dalam melakukan klasifikasi, karena secara konsep model tersebut dibangun dengan memperhatikan hubungan atau dependensi diantara pasangan fitur-fitur objek (Irwan, 2007, 2011, 2012). Sedangkan *Maximum Likelihood* (ML) adalah salah satu *estimator* yang telah banyak digunakan

dan telah teruji penggunaannya dalam banyak aspek (Duda,1973).Secara teori

dalam mengenali suatu objek tidak boleh lepas dari dua hal yaitu metode dalam membangun model pengenalan serta *estimator* model yang digunakan, akan tetapi pada kenyataannya hanya memperhatikan dua hal tersebut ternyata masih belum cukup untuk menghasilkan suatu model pengenalan terbaik yang memberikan tingkat akurasi tinggi dalam pengenalan. Faktor lain yang dimungkinkan memberi kontribusi dalam menghasilkan model terbaik adalah kondisi dari dimensi atau fitur objek (*image*), khususnya fitur objek untuk pembelajaran (*training*) dalam membangun model pengenalan. Bertolak dari dimensi atau fitur objek tersebut, muncul permasalahan bagaimana mendapat kan atau membangun model pengenalan terbaik dengan *Tree-Augmented Network* (TAN) dan *estimator*

Maximum Likelihood (ML) berdasarkan dimensi atau fitur objek *training*.

1. PRINSIP *MAXIMUM LIKELIHOOD* (ML)

Bila diketahui data sampel *training* D yang terdiri dari n sampel x_1, x_2, \dots, x_n dengan asumsi identik, independen distribusi variabel random (i.i.d), dapat ditentukan fungsi *likelihood*

$$p(D|\theta) = \prod_{k=1}^n p(x_k|\theta)$$

(1)

dengan θ adalah parameter distribusi, maka prinsip dari *estimator Maximum Likelihood* (ML) untuk mendapatkan taksiran parameter θ adalah dengan memaksimumkan fungsi $p(D|\theta)$ (Duda,1973).

3. BAYESIAN NETWORKS

Bila diketahui $U = \{X_1, \dots, X_n\}$ adalah himpunan variabel random diskrit, maka *bayesian network* dinotasikan sebagai *Directed Acyclic Graph* (DAG) yang merupakan join distribusi probabilitas U . Secara formal join distribusi probabilitas tersebut merupakan sebuah pasangan *Directed Acyclic Graph* (G) dengan parameter *network* (Θ) yang ditulis $B = \langle G, \Theta \rangle$. Parameter *network* (Θ) berisi suatu parameter $\theta_{x_i|\Pi_{x_i}} = P_B(x_i | \Pi_{x_i})$ untuk setiap nilai $x_i \in X_i$ dan $\Pi_{x_i} \in \Pi_{X_i}$, dimana Π_{X_i} merupakan sekumpulan *parent* dari X_i didalam G . Sehingga *bayesian network* didefinisikan sebagai sebuah *uniqu* join distribusi probabilitas U yang unik (Friedman,1996,1997), sebagai berikut:

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \Pi_{X_i}) = \prod_{i=1}^n \theta_{(x_i|\Pi_{x_i})}$$

(2)

3.1 *Tree-Augmented Network* (TAN)

Pembahasan *Tree-Augmented Network* (TAN) berkaitan dengan *naïve Bayes* klasifier karena *Tree-Augmented*

merupakan perpanjangan atau kelanjutan dari *naïve Bayes* klasifier. *Naïve Bayes* Klasifier diperoleh dengan pembelajaran data training D yaitu dengan menentukan kodisional probabilitas pada setiap atribut X_i bila diberikan variabel kelas C . Karenanya *naïve Bayes* tidak realistis untuk diterapkan pada data riil, sehingga terdapat perbaikan pada *naïve Bayes* yang disebut *augmented naïve Bayes*. Membangun *augmented naïve Bayes classifier* equivalen dengan mencari *Bayesian Network* yang baik dengan variabel kelas C sebagai *root* (Friedman,1997). Karena komputasi yang intensif maka solusi yang efisien dalam menemukan *Bayesian network* adalah diperbolehkannya saling mempengaruhi diantara variabel. Selanjutnya network yang dihasilkan disebut *Tree-Augmented Network* (TAN).

3.2 *Construct-TAN*

Construct-TAN (Amy, 2005) (Friedman, 1996, 1997) (Murphy,2001) dibangun berdasarkan prosedur Chow dan Liu dengan sedikit modifikasi. Prosedur *Construct-TAN* ada lima tahap, sebagai berikut:

1. Menghitung $I_{\hat{p}_D}(X_i, X_j | C)$ diantara setiap pasangan atribut, $i \neq j$.
2. Membangun *graph* komplit tak berarah dengan *node*-nya merupakan atribut X_1, \dots, X_n . Sedangkan bobot pada *edge* $X_i X_j$ adalah $I_{\hat{p}_D}(X_i, X_j | C)$.
3. Membangun *maximum weighted spanning tree* (MWST) dengan algoritma Prim's (Levitin,2003).
4. Transformasi hasil dari *tree* tak berarah menjadi sebuah *tree* berarah dengan memilih sebuah *root* variabel dan mensetting arah *edge* dari *root* variabel tersebut.
5. Bangun model TAN dengan menambah simpul (*vertex*) yang diberi

label C dan menambah *edge* atau *arc* dari C ke setiap X_i .

Untuk bobot $(I_{\hat{P}_D}(X_i, X_j | C))$ diperoleh dengan menggunakan fungsi *conditional mutual information*, seperti pada persamaan 3 (Duda,1973) (Friedman, 1996, 1997).

$$I_{P_D}(X_i, X_j | C) = \sum_{x \in X_i, y \in Y} P(X_i, X_j, c) \log \frac{P(X_i, X_j | c)}{P(X_i | c)P(X_j | c)}$$

(3)

Untuk variabel kontinu, *mutual information* antara dua variabel kontinu dianggap sebagai *joint distribution (bivariate normal)* bila diberikan C (C =kelas). Fungsi *joint distribution* X_i dan X_j dengan syarat $C=c$ mengikuti distribusi *bivariate normal* dengan mean $\mu_{ij|c}$ dan *covariant* $\sum_{ij|c}$ (Michael,2006). Sehingga nilai *mutual information* antara variabel X_i dan X_j dengan syarat C dapat ditentukan dengan mengacu pada persamaan 4 (Aritz,2006).

$$I(X_i, X_j | C) = -\frac{1}{2} \sum_{c=1}^r P(c) \log(1 - \rho_c^2(X_i, X_j))$$

(4)

$$\rho_c(X_i, X_j) = \frac{\sigma_{ij|c}}{\sqrt{\sigma_{i|c}^2 \sigma_{j|c}^2}}$$

dimana

(5)

adalah koefisien korelasi antara X_i dan X_j dengan syarat $C=c$. Adapun tahapan-tahapan dalam *Construct-TAN* prinsipnya sama dengan *Construct-TAN* untuk variabel diskret.

3.3 Estimasi Parameter Model TAN

Dengan mengacu pada sub-bab sebelumnya dan diasumsikan bahwa struktur model tetap maka nilai parameter Θ dapat ditentukan untuk setiap kelasnya, seperti

pada persamaan 6 (Friedman, 1996,1997) (Jesus, 1999)

$$\theta_{x_i | \Pi_{x_i} | C} = \hat{P}_D(x_i | \Pi_{x_i} | C) = \frac{\hat{P}_D(x_i, \Pi_{x_i} | C)}{\hat{P}_D(\Pi_{x_i} | C)}$$

(6)

dengan $\hat{P}_D(x_i, \Pi_{x_i} | C) \approx N(\mu_{x_i | \Pi_{x_i} | C}, \Sigma_{x_i | \Pi_{x_i} | C})$ dan $\hat{P}_D(\Pi_{x_i} | C) \approx N(\mu_{\Pi_{x_i} | C}, \Sigma_{\Pi_{x_i} | C})$.

4. METODE PENELITIAN

4.1. Construct-TAN dengan ML

Langkah-langkah membangun Model TAN dengan menggunakan RLME untuk pengenalan suatu objek citra/*image* dapat dilihat secara detail pada Algoritma Membangun Model TAN dengan ML.

Algoritma Membangun Model TAN dengan *Maximum Likelihood* (ML)

Construct – TAN(D)

for setiap X_i, X_j do

Tentukan w_{ij} (bobot)

{ $w_{ij} = I(X_i, X_j | C)$ berdasarkan parameter $\sum_{ij|c}$ yang diperoleh dengan ML }

$UG \leftarrow$ BangunGraphTakBerarah(w)

$G \leftarrow$ MWST(UG)

{maximum weighted spanning tree}

$T \leftarrow$ BangunGraphBerarah($G, root$)

$TAN \leftarrow$ TambahKelas(T)

return TAN

Prosedur *Construct-TAN* adalah prosedur yang digunakan dalam *learning TAN* dan mengacu pada persamaan 6 dalam menentukan parameter atau bobot model.

4.2. Eksperimen

Untuk mendapatkan model pengenalan TAN terbaik terhadap objek tertentu dengan bantuan *estimator ML* untuk menghitung parameter modelnya, dicoba melakukan beberapa *training* terhadap data dengan objek sama, dan dimensi atau fitur

objek yang berbeda-beda. Hasil *training* terhadap objek *training* yang sama dengan dimensi atau fitur objek yang berbeda-beda kemudian dibandingkan tingkat akurasi sistemnya dalam pengenalan.

Data yang digunakan dalam eksperimen adalah data nyata dalam bentuk *image* yang terdiri dari 5 kelas (objek) yaitu buah apel, boneka, tekstur tanah, kaleng, dan sepatu seperti pada Gambar 1. yang masing-masing kelas memiliki sampel berukuran 15 pengamatan (Irwan, 2007).



Gambar 1. Data nyata dengan lima jenis objek (buah apel, boneka, tekstur tanah, kaleng, dan sepatu)

Sedang skenario data percobaannya dapat dilihat pada Tabel 1 yaitu skenario data percobaan dengan beberapa dimensi atau fitur objek.

Tabel 1. Skenario data percobaan dengan beberapa dimensi atau fitur objek.

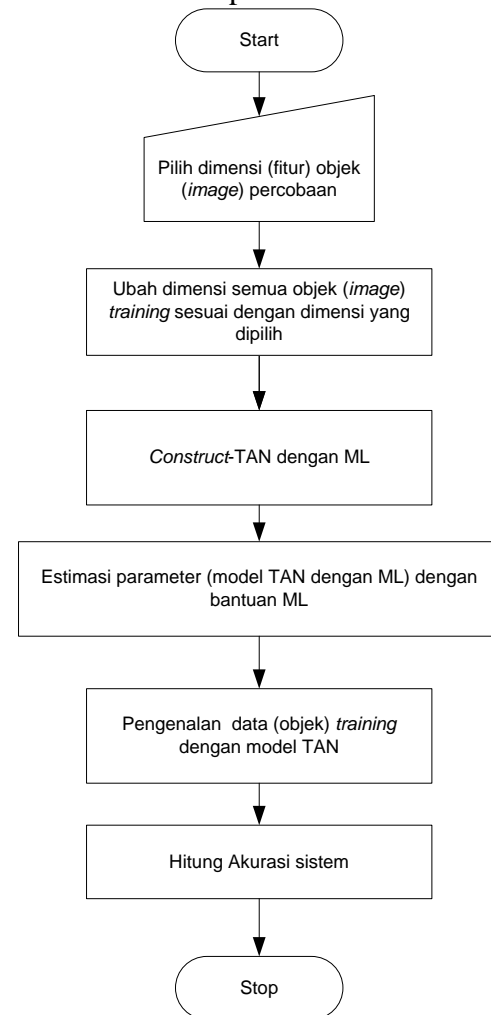
Percobaan	Dimensi Objek	Ukuran Fitur
1	2x2	4
2	3x2	6
3	3x3	9
4	4x3	12
5	4x4	16
6	5x4	20
7	5x5	25
8	6x5	30
9	6x6	36
10	7x6	42
11	7x7	49
12	8x7	56
13	8x8	64
14	9x8	72
15	9x9	81
16	10x9	90
17	10x10	100

Maksud dimensi atau fitur objek pada setiap percobaan pada Tabel 1 adalah lima

jenis objek (buah apel, boneka, tekstur tanah, kaleng, dan sepatu) dalam bentuk *image* dengan setiap percobaan ukuran dimensi atau fiturnya sama. Sebagai misal percobaan pertama dimensi objek 2x2 atau dengan 4 fitur, maka dimensi *image* untuk lima jenis objek tersebut dibuat sama dengan ukuran 2x2 atau sebanyak 4 fitur.

4.3 Pelaksanaan Eksperimen

Pelaksanaan eksperimen (uji coba) terhadap data nyata yang meliputi lima jenis objek atau kelas dalam bentuk *image*, dilakukan dengan skenario data percobaan pada Tabel 1. Sedangkan pelaksanaan eksperimen dengan menggunakan data skenario tersebut dapat dilihat langkah-langkahnya secara lengkap pada Gambar 2. Pelaksanaan Eksperimen.



Gambar 2. Pelaksanaan Eksperimen

Pada langkah pertama, setiap percobaan pada Tabel 1, dengan dimensi objek atau fitur yang berbeda-beda akan dibangun model TAN dengan bantuan *estimator* ML. Kemudian dilakukan estimasi parameter model TAN dengan bantuan *estimator* ML juga. Setelah dihasilkan model TAN beserta parameter modelnya, langkah berikutnya adalah melakukan pengenalan objek *training* dan menghitung akurasi sistem. Dari hasil perhitungan akurasi sistem pada setiap percobaan, selanjutnya dipilih model TAN dengan tingkat akurasi sistem tertinggi yang kemudian dianggap sebagai model terbaik dalam pengenalan.

5. HASIL DAN PEMBAHASAN

Hasil percobaan dengan menggunakan skenario data pada Tabel 1, selengkapnya dapat dilihat pada Tabel 2

Tabel 2. Tingkat Akurasi Model TAN Dalam Mengenali Objek Training

Dimensi Objek	Ukuran Fitur	Akurasi (%)	Kesalahan (%)
2x2	4	82.6667	17.3333
3x2	6	78.6667	21.3333
3x3	9	96	4
4x3	12	98.6667	1.3333
4x4	16	98.6667	1.3333
5x4	20	98.6667	1.3333
5x5	25	100	0
6x5	30	98.6667	1.3333
6x6	36	100	0
7x6	42	100	0
7x7	49	98.6667	1.3333
8x7	56	97.3333	2.6667
8X8	64	97.3333	2.6667
9X8	72	94.6667	5.3333
9X9	81	77.3333	22.6667
10x9	90	58.6667	41.3333
10x10	100	61.3333	38.6667

Model TAN hasil *training* dari beberapa dimensi atau fitur objek (*image*), memperlihatkan bahwa untuk dimensi atau fitur yang kecil tingkat akurasi sistem

model TAN dalam melakukan pengenalan relatif lebih rendah. Rendahnya tingkat akurasi sistem dibandingkan dengan yang lainnya, ditunjukkan oleh dimensi objek 2x2 (4 fitur) sebesar 82, 6667%, dimensi 3x2 (6 fitur) 78,6667% . Demikian halnya untuk dimensi atau fitur yang besar, model TAN yang dihasilkan memberikan tingkat akurasi sistem juga relatif lebih rendah dibandingkan lainnya. Rendahnya tingkat akurasi tersebut ditunjukkan pada dimensi 9x9 (81 fitur) sebesar 77,3333%, dimensi 10x9 (90 fitur) sebesar 58,6667% dan dimensi 10x10 (100 fitur) sebesar 61,3333%.

Tingkat akurasi paling tinggi dari eksperimen, dihasilkan oleh model TAN pada dimensi 5x5 (25 fitur), dimensi 6x6 (36 fitur) dan 7x6(42 fitur) masing-masing dengan tingkat akurasi 100%. Hasil ini menunjukkan bahwa pada dimensi tersebut model TAN yang dihasilkan merupakan model pengenalan terbaik. Hal penting lain yang dapat diambil dari hasil eksperimen tersebut adalah semakin kecil dimensi atau semakin besar dimensi objek yang digunakan akan berdampak model TAN yang dihasilkan semakin kurang baik. Kurang baiknya model TAN pada dimensi atau fitur objek yang semakin kecil bisa jadi karena semakin kecil dimensi atau fitur banyak informasi penting dari objek yang hilang. Sedangkan kurang baiknya model TAN yang dihasilkan pada dimensi yang semakin besar karena semakin besar dimensi atau fitur objek menyebabkan terjadinya informasi *overflow*.

6. KESIMPULAN

Kehandalan model serta *estimator* yang digunakan pada kenyataan belum cukup untuk menghasilkan model pengenalan terbaik, akan tetapi ada faktor lain yang member kontribusi besar yaitu kondisi dari dimensi atau fitur objek yang digunakan dalam membangun model tersebut. Hasil eksperimen menunjukkan untuk dimensi atau fitur objek semakin kecil atau untuk dimensi atau fitur objek yang semakin

besar menghasilkan model TAN dengan tingkat akurasi sistemnya relatif lebih rendah. Kondisi tersebut menunjukkan pada dimensi atau fitur yang semakin kecil banyak informasi penting yang hilang, begitu pula semakin besar dimensi atau fitur objek yang digunakan akan menyebabkan informasi *overflow* sehingga menghasilkan model pengenalan yang kurang baik.

8. REFERENSI

1. Amy Ratnakaran, "Bayesian Network", Applied Statistics Honours, Department of Mathematics and Statistics, University of Melbourne, 2005.
2. Aritz Pérez, Pedro Larrañaga, and Inaki Inza, "Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive bayes", International Journal of Approximate Reasoning, vol.43.no.1,hal 1–25, 2006
3. R. Duda dan P. Hart, "Pattern Classification and Scene Analysis", John Wiley and Sons, Inc., New-York, USA, 1973.
4. Friedman, N. and Goldszmidt, M, "Building Classifiers Using Bayesian Networks", Thirteenth National Conference on Artificial Intelligence, 1996.
5. Friedman, N. D. Geiger, and M. Goldszmidt, "Bayesian network classifiers", Machine Learning, vol.29, hal 131–163, 1997.
6. Irwan B.S., "Pengenalan Objek dengan
7. Regularized Latent Maximum Entropy (RLME) dan Tree-Augmented Network (TAN)", Master Tesis, Institut Teknologi Sepuluh Nopember Surabaya. 2007
8. Irwan B.S., "Pengenalan Objek dengan
9. RLME dan TAN (Perbandingan RLME_ML dan ML_ML)", Makalah Seminar Nasional Pascasarjana VII, Institut Teknologi Sepuluh Nopember Surabaya. 2007
10. Irwan B.S, "Perbandingan Metode RLME dan ML Dalam Membangun Model TAN Untuk Mengenali Objek Dengan Fitur Berdistribusi Normal", MATICS (jurnal Ilmu Komputer dan Teknologi Informasi UIN Malang), vol.4. no.4, hal 148-154, September 2011.
11. Irwan B.S., "Perbandingan Struktur Model
12. TAN dengan Estimator RLME Dan ML Untuk Mengenali Objek Dengan Fitur Berdistribusi Normal (Studi Kasus Data Sintesis)", Prosiding Seminar Nasional SciETec Program Magister dan Doktor Fakultas Teknik, Universitas Brawijaya, 2012.
13. Jesus Cerquides, "Applying General Bayesian Techniques to Improve TAN Induction", UBS AG Bahnhofstrasse 45, 1999.
14. Levitin, "Introduction The Design & Analysis of Algorithms", Villanova University, 2003.
15. Michael P. Ashley-Rollman, Lucia Castellanos Pérez-Bolde, "TAN

