

ALGORITMA BAYESIAN NETWORK UNTUK SIMULASI PREDIKSI PEMENANG PILKADA MENGGUNAKAN MSBNX

Andi Lukman *)
Muh Nadzirin Anshari Nur **)

Abstract : *Abstract-Quick count is understood as the process of counting quickly and appropriately in an implementation ELECTION. In addition to Quick count, another thing that is done to determine the candidate's electability of candidates is to conduct an initial survey before the election takes place, the survey can be carried out by a competent agency survey. So that an accurate survey data generated, it takes the data from many provincial, district / city. It requires the help of data mining methods in particular Bayesian Network, because the data has a causal relationship to one another either directly or indirectly to determine the winner ELECTION. Many factors or variables which affect support for the candidates, there are 8 variables in the study, consisting of 7 predictor variables are: Position Status, Conformity Religion, Ethnic Equality, Popularity, Specialization, and Support Support Survey I Survey II and the response variable is 1 Status Quick Count. The above factors in predicting the winner was simulated to use the Bayesian Network tool is Microsoft Bayesian Network Editor (MSBNx). There are 4 stages of data analysis, namely: define variable and related categories in the network, making causality diagram, filling the conditional probability each variable, and include evidence to see the results of prediction. Minimum criteria for a successful candidate in the quick count result is above 50% chance of success, it was found that the variable Support Survey I and II are very influential in winning ELECTION, where if two variables do not have evidence or not successful, then the chances of a quick count would not succeed .*

Keywords: *component; Simulation; Bayesian Network; Prediction Winner ELECTION; MSBNx.*

PENDAHULUAN

Pada setiap pelaksanaan Pilkada berbagai elemen masyarakat berlomba-lomba untuk mendapatkan informasi tentang hasil akhir PILKADA. Berbagai metode penghitungan dikolaborasikan untuk mencapai tujuan itu, Sejauh ini sebuah metode yang populer telah diterapkan di Indonesia yaitu *quick count*. *Quick count* dipahami sebagai proses penghitungan secara cepat dan tepat dalam sebuah pelaksanaan PILKADA.

Penghitungan cepat dapat dilakukan dengan berbagai cara maupun alat yang digunakan. Cara paling aman dan sudah pasti dapat diterima oleh semua lapisan masyarakat tanpa perlu penjelasan tambahan adalah memantau semua TPS dan menghitung perolehan masing-masing kandidat, lalu kemudian menampilkannya dalam sebuah grafik maupun angka

persentase, ini biasa disebut *Paralel Vote Tabulation* (PVT).

Selain *quick count*, hal lain yang dilakukan kandidat untuk mengetahui Elektabilitas calon adalah dengan melakukan survey awal sebelum pilkada berlangsung, survey dapat dilakukan oleh lembaga survey independent ataupun tim sukses kandidat sendiri. Banyak Faktor yang mempengaruhi dukungan terhadap para kandidat termasuk variable yang fluktuatif antara survey I maupun survey II bisa berubah secara dinamis, termasuk faktor faktor *Incumbent*, Agama, Etnis dan lain sebagainya [4]. Untuk keakuratan data, dibutuhkan data yang banyak dari berbagai propensi, kabupaten/kota. Hal ini membutuhkan *Data Mining*, khususnya metode *Bayesian Network* karena setiap variable mempunyai hubungan kausalitas, baik secara langsung maupun tak langsung. Untuk mensimulasikan prediksi

pemenang PILKADA dapat menggunakan MSBNx (Microsoft Bayesian Network) sebagai salah satu *Tools Bayesian Network*.

Tujuan yang ingin dicapai dari penelitian ini adalah menerapkan Algoritma Bayesian Networks untuk Simulasi Prediksi Pemenang PILKADA menggunakan MSBNx dan melihat sejauh mana Faktor-faktor yang mempengaruhi hasil PILKADA berdasarkan simulasi yang dibuat.

TINJAUAN PUSTAKA

A. Teori Bayes

Teori Bayesian diadopsi dari nama penemunya yaitu Thomas Bayes sekitar tahun 1950, yang sering ditemukan pada studi-studi ilmu statistika yang berbasis pada teorema atau aturan Bayes. Teori Bayesian merupakan sebuah teori kondisi probabilitas yang memperhitungkan probabilitas suatu kejadian (hipotesis) bergantung pada kejadian lain (bukti) [6]. Pada dasarnya, teorema tersebut mengatakan bahwa kejadian dimasa depan dapat diprediksi dengan syarat kejadian sebelumnya telah terjadi.

Statement teori bayes:

$$P(A|B)P(B) = P(A,B) = P(B|A)P(A) \dots\dots\dots(1)$$

dimana $P(A/B)$ adalah probabilitas gabungan kejadian A dan B . Membagi kedua sisi dengan $P(B)$, didapat:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots\dots\dots(2)$$

Secara umum teorema Bayes dapat dituliskan dalam bentuk:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)} \dots\dots\dots(3)$$

Jika $\{A_i\}$ membentuk partisi dari ruang kejadian, untuk setiap A_i dalam partisi.

B. Bayesian Network

Beberapa istilah yang biasa digunakan untuk menyebutkan *Bayesian Network*, antara lain : *Bayesian Belief*, *Bayesian Belief Network* dan *Belief Network*. *Bayesian network* merupakan salah satu *Probabilistic Graphical Model* (PGM) yang sederhana yang dibangun dari teori probabilistik dan teori graf. Teori probabilitas berhubungan langsung dengan data sedangkan teori graf berhu-

bungan langsung dengan bentuk representasi yang ingin didapatkan. Pengetahuan tersebut direpresentasikan secara kualitatif menggunakan struktur graf dan secara kuantitatif menggunakan parameter-parameter numeric [2]. Sebagai contoh, sebuah *bayesian network* dapat mewakili hubungan probabilistik antara penyakit dan gejala. *Bayesian network* dapat digunakan untuk menghitung probabilitas dari kehadiran berbagai gejala penyakit.

Metode *bayesian network* merupakan metode yang baik di dalam *machine learning* berdasarkan data *training*, dengan menggunakan probabilitas bersyarat sebagai dasarnya. Metode ini terdiri dari 2 bagian utama, yaitu :

1. *Directed Acyclic Graph* (DAG)

Struktur graf *bayesian network* disebut dengan *Directed Acyclic Graph* (DAG) yaitu graf berarah tanpa siklus berarah. DAG terdiri dari *node* dan *edge*. *Node* merepresentasikan variabel acak dan *edge* merepresentasikan adanya hubungan ketergantungan langsung dan dapat juga diinterpretasikan sebagai pengaruh (sebab-akibat) langsung antara variabel yang dihubungkannya. Tidak adanya *edge* menandakan adanya hubungan kebebasan kondisional di antara variabel.

Struktur grafis *bayesian network* ini digunakan untuk mewakili pengetahuan tentang sebuah domain yang tidak pasti. Secara khusus, setiap node dalam grafik merupakan variabel acak, sedangkan ujung antara *node* mewakili probabilistik yang bergantung di antara variabel-variabel acak yang sesuai. Kondisi ketergantungan ini dalam grafik sering diperkirakan dengan menggunakan statistik yang dikenal dengan metode komputasi. Oleh karena itu, *bayesian network* menggabungkan prinsip-prinsip dari teori graf, teori probabilitas, ilmu pengetahuan komputer, dan statistik.

2. Himpunan Parameter

Himpunan parameter mendefinisikan distribusi probabilitas kondisional untuk setiap variabel. Pada *bayesian network*, *nodes* berkorespondensi dengan variabel acak. Tiap *node* diasosiasikan dengan sekumpulan peluang bersyarat, $p(x_i|A_i)$ sehingga x_i adalah variabel yang diasosiasikan dengan *node* dan A_i adalah set dari *parent* dalam graf.

C. Microsoft Bayesian Network Editor (MSBNx)

Microsoft Bayesian Network Editor (MSBNx) merupakan salah satu Tool yang digunakan untuk melakukan analisa *Bayesian Network*. MSBNx adalah *tool* berbasis *Windows* untuk membuat, simulasi, dan mengevaluasi *Bayesian Networks* yang dibuat oleh *Microsoft Research* [1]. Modul instalasi aplikasi termasuk file *help* lengkap dan contoh jaringan, *Bayesian Network* dikodekan dalam format file XML.

Beberapa tahap dalam membuat Bayesian Network menggunakan MSBNx *tools*, yaitu:

1. Langkah pertama yaitu membuka aplikasi MSBNx dan memilih *File, New Model*.
2. Kemudian membuat Node-node atau variable-variable yang dibutuhkan oleh Bayesian network dengan cara klik kanan pada layar kemudian pilih *Add Node*, kemudian mengisi nama variable.
3. Setelah semua node dibuat, langkah selanjutnya membuat hubungan kausalitas antar node, dengan menklik 2 node yang berhubungan (sambil tekan tombol CTRL) kemudian klik kanan dan pilih *Add Dependency Arc*.
4. Langkah selanjutnya memberikan nama state dan property dari setiap variable dengan mengklik kanan node kemudian memilih *States & Properties*, kemudian mengisinya.
5. Selanjutnya mengisi nilai probabilitas setiap state pada variable dengan cara

klik kanan pada node kemudian pilih *assess*.

6. Untuk menampilkan nilai dari setiap variable, klik kanan pada node kemudian pilih *Bar Chart*.
7. Sampai langkah 6, Bayesian network berdasarkan historical data telah jadi. Hal ini digunakan untuk memprediksi dengan cara memasukkan *evidence*. Evidence merupakan hasil pengamatan lapangan terhadap salah satu atau beberapa variable. Untuk memasukkan evidence, klik kanan node yang mempunyai evidence kemudian masukkan nilai evidencenya.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data simulasi, diasumsikan bahwa data berasal dari penelitian PILKADA dari berbagai propinsi dan kabupaten/kota. Penelitian harus dilakukan oleh lembaga survey independen yang mempunyai kompetensi. Setelah melakukan penelitian, diasumsikan bahwa historical data telah siap diolah untuk membantu pengambilan keputusan. Karena data yang diperoleh sangat banyak dan kompleks, untuk memudahkan dalam analisis, maka digunakan *data mining* dengan metode *Bayesian Network*.

Untuk memudahkan dalam analisis data Bayesian network, digunakan tool MSBNx. Ada 4 tahap analisis data yaitu : menentukan variable dan kategori yang terkait dalam jaringan, membuat diagram kausalitas, mengisi *conditional probability* setiap variable, dan memasukkan evidence untuk melihat hasil prediksi.

D. Variable dan Kategori

Tahap pertama adalah menentukan variable beserta kategori yang mempengaruhi prediksi pemenang PILKADA. Data terdiri dari 8 variable, diantaranya 7 variable predictor dan 1 variable respon. Hasil pendiskretan 8 variable tersebut sehingga setiap variable memiliki 2 kategori yang dapat dilihat pada Tabel 1.

Tabel 1 Daftar Variabel dan Kategori

No.	Variable	Kategori
1	Status Jabatan	Incumbent (1) Tidak Incumbent (0)
2	Kesamaan etnis	Sama (1) Tidak Sama (0)
3	Kesesuaian agama	Sesuai (1) Tidak Sesuai (0)
4	Dukungan Survey II Pra quick count	Sukses (1) Gagal (0)
5	Dukungan Survey I Pra quick count	Sukses (1) Gagal (0)
6	Propularitas saat survey I Pra quick count	Populer (1) Tidak Populer (0)
7	Peminatan saat survey Survey I Pra quick count	Diminati (1) Tidak Diminati (0)
8	Status quick count	Sukses (1) Gagal (0)

Diagram Kausalitas

Tahap kedua adalah membuat diagram kausalitas atau *causal network*, yaitu hubungan sebab akibat setiap variable baik secara langsung maupun tak langsung. Gambar diagram kausalitas prediksi pemenang PILKADA dapat dilihat pada Gambar 1.

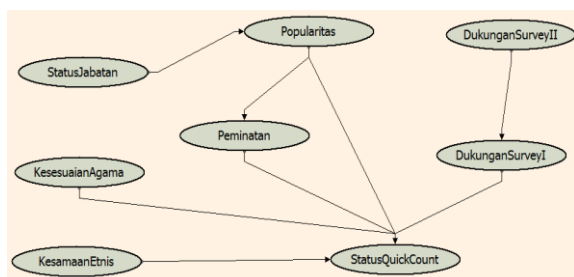


Figure 1. Diagram Kausalitas prediksi Pemenang PILKADA

Variable Kesamaan Etnis dan Kesesuaian Agama berhubungan langsung dengan StatusQuickCount. Sedangkan variable StatusJabatan menjadi *node parent* bagi variable Popularitas. Popularitas bersifat bebas bersyarat menjadi *node parent* bagi variable Peminatan yang berdasarkan asumsi bahwa seseorang dikatakan populer karena *track recordnya* sebagai *incumbent* (StatusJabatan), sehingga masyarakat mempunyai minat

(Peminatan) yang kuat yang berimplikasi pada StatusQuickCount.

Variable DukunganSurveyII dan Dukungan SurveyI merupakan variable berseri yang juga mempengaruhi Status Quick Count. Variable Status Quick Count merupakan variable output yang akan dicari dalam penelitian ini, variable ini merepresentasikan dukungan masyarakat dalam PILKADA.

E. Conditional Probability

Tabulasi probabilitas setiap variable atau *Conditional Probability* menggunakan data simulasi. Adapun data probabilitas dapat dilihat pada Tabel 2 sampai dengan Tabel 9. Pada setiap tabel diisi dengan nilai antara 0 sampai 1, nilai 1 menyatakan persentase 100%, dengan kata lain, jika ingin melihat nilai dalam bentuk persentase, nilai dikalikan dengan 100.


TABLE I. NILAI VARIABLE STATUS JABATAN

StatusJabatan		
Incumbent	Not Incumbent	bar charts
0.25	0.75	

Tabel 2 menunjukkan persentase probabilitas kandidat kepala daerah yang merupakan Incumbent sebesar 25 % dan


bukan Incumbent sebesar 75%. Selengkapnya dapat dilihat pada Tabel 2.

TABLE II. NILAI VARIABLE KESAMAAN ETNIS

KesamaanEtnis		
Sama	Tidak Sama	bar charts
0.9	0.1	

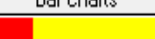
Dari data yang diperoleh (simulasi), didapatkan bahwa 90% kandidat kepala daerah mempunyai kesamaan etnis dengan pemilih. Selebihnya (10%) tidak sama. Selengkapnya dapat dilihat pada Tabel 3.

TABLE III. NILAI VARIABLE KESESUAIAN AGAMA

KesesuaianAgama		
Sesuai	Tidak Sesuai	bar charts
0.95	0.05	



Sementara kesesuaian agama kandidat dengan pemilih mempunyai persentase 95%, hanya 5% yang tidak sesuai. Selengkapnya dapat dilihat pada Tabel 4.

TABLE IV. NILAI VARIABLE DUKUNGAN SURVEY II PRA QUICK COUNT

DukunganSurveyII		
Sukses	Gagal	bar charts
0.25	0.75	

Survey Dukungan II menunjukkan 25% kandidat sukses, sementara 75% tidak. Selengkapnya dapat dilihat pada Tabel 5.

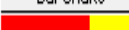

TABLE V. NILAI VARIABLE DUKUNGAN SURVEY I PRA QUICK COUNT

Parent Node(s)	DukunganSurveyI		bar charts
DukunganSurveyII	Sukses	Gagal	
Sukses	0.7	0.3	
Gagal	0.3	0.7	

Survey dukungan I dipengaruhi oleh variable survey dukungan II (sebagai

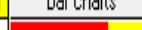

parentnya), sehingga dalam mengisi nilai tabulasi persentasenya harus berlandaskan setiap criteria dari *parentnya*. Nilai-nilai variable dapat dilihat pada Tabel 6.

TABLE VI. NILAI VARIABLE POPULARITAS SAAT SURVEY I PRA QUICK COUNT

Parent Node(s)	Popularitas		bar charts
Status.Jabatan	Populer	Tidak Populer	
Incumbent	0.65	0.35	
Not Incumbent	0.35	0.65	

Popularitas saat survey I pra quick count dipengaruhi oleh parent nodenya yaitu Status Jabatan, sehingga nilainya bergantung pada *parentnya*. Nilai-nilai variable dapat dilihat pada Tabel 7.

TABLE VII. NILAI VARIABLE PEMINATAN SAAT SURVEY I PRA QUICK COUNT

Parent Node(s)	Peminatan		bar charts
Popularitas	Diminati	Tidak Diminati	
Populer	0.7	0.3	
Tidak Populer	0.3	0.7	

Variable peminatan saat survey I pra quick count dipengaruhi oleh parent node Polularitas, sehingga setiap nilainya bergantung pada variable probabilitas. Nilai-nilai variable dapat dilihat pada Tabel 8.

Secara keseluruhan, variable status quick count dipengaruhi oleh seluruh variable, baik secara langsung maupun tak langsung. Nilai variable ini diisi menurut variable-variable yang secara langsung mempengaruhinya. Nilai-nilai variable dapat dilihat pada Tabel 9.

TABLE VIII. NILAI VARIABLE STATUS QUICK COUNT

Parent Node(s)					StatusQuickCount		
KesesuaianAgama	KesamaanEtnis	Popularitas	Peminatan	DukunganSurveyI	Sukses	Gagal	bar charts
Sesuai	Sama	Populer	Diminati	Sukses	0.8	0.2	
				Gagal	0.2	0.8	
			Tidak Diminati	Sukses	0.6	0.4	
		Gagal		0.4	0.6		
		Tidak Populer	Diminati	Sukses	0.85	0.15	
				Gagal	0.15	0.85	
	Tidak Diminati		Sukses	0.7	0.3		
		Gagal	0.3	0.7			
	Tidak Sama	Populer	Diminati	Sukses	0.75	0.25	
				Gagal	0.25	0.75	
			Tidak Diminati	Sukses	0.55	0.45	
		Gagal		0.45	0.55		
Tidak Populer		Diminati	Sukses	0.5	0.5		
			Gagal	0.5	0.5		
	Tidak Diminati	Sukses	0.4	0.6			
Gagal		0.6	0.4				
Tidak Sesuai	Sama	Populer	Diminati	Sukses	0.75	0.25	
				Gagal	0.25	0.75	
			Tidak Diminati	Sukses	0.65	0.35	
		Gagal		0.35	0.65		
		Tidak Populer	Diminati	Sukses	0.6	0.4	
				Gagal	0.4	0.6	
	Tidak Diminati		Sukses	0.55	0.45		
		Gagal	0.45	0.55			
	Tidak Sama	Populer	Diminati	Sukses	0.4	0.6	
				Gagal	0.6	0.4	
			Tidak Diminati	Sukses	0.45	0.55	
		Gagal		0.55	0.45		
Tidak Populer		Diminati	Sukses	0.4	0.6		
			Gagal	0.6	0.4		
	Tidak Diminati	Sukses	0.3	0.7			
Gagal		0.7	0.3				

Setelah pengisian nilai setiap variable, kita dapat menampilkan diagram kausalitas lengkap dengan nilai probabilitas setiap variable. Diagram inilah yang dijadikan dasar dari prediksi jika ditemukan *evidence* salah satu atau beberapa variable sesuai dengan fakta survey baru. Diagram yang dilengkapi nilai probabilitas setiap variable dapat dilihat pada gambar 2.

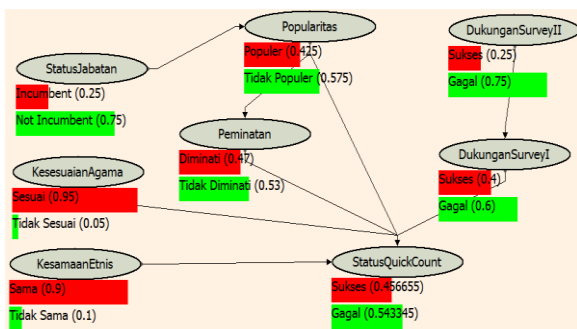


Figure 2. Diagram Kausalitas Dengan Nilai Probabilitas Setiap Variable

F. Evidence

Evidence atau fakta dunia nyata merupakan hasil survey saat ini berdasarkan criteria kandidat kepala

daerah. Bayesian Network digunakan disini karena kecenderungan tidak semua variable mempunyai *Evidence*, namun dengan hanya menggunakan satu atau beberapa *evidence*, Bayesian Network dalam memprediksi peluang Kandidat untuk memenangkan PILKADA dengan bantuan *historical data mining*.

Kriteria minimum seorang kandidat untuk sukses dalam *quick count* adalah menghasilkan peluang sukses diatas 50%. Pada penelitian ini ditemukan beberapa criteria yang memungkinkan peluang seorang kandidat memenangkan PILKADA. Beberapa diantaranya sebagai berikut:

- Jika ditemukan *evidence* dukungan survey II pra quick count adalah sukses(1), maka pada survey selanjutnya akan meningkat dari 40% menjadi 70% dan kondisi variable yang lain tidak diketahui, maka kondisi ini akan meningkatkan peluang sukses *quick count* dari 45,66% menjadi 58.66% Diagram

kausalitasnya dapat dilihat pada Gambar 3.

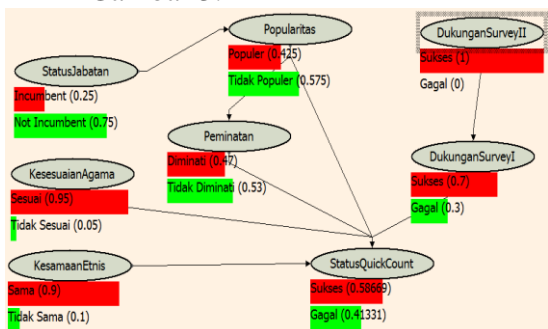


Figure 3. Diagram Kausalitas Dengan Evidence DukunganSurveyII Sukses(1)

- Jika dukungan survey I pra quick count adalah sukses(1) dan kondisi variable lainnya tidak diketahui, maka peluang sukses *quick count* PILKADA akan meningkat dari 45.66% menjadi 71.67%. Hal ini dapat dilihat pada Gambar 4.

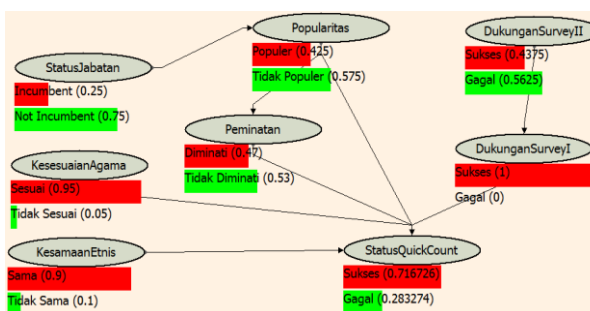


Figure 4. Diagram Kausalitas Dengan Evidence DukunganSurveyI Sukses(1)

- Jika Kandidat populer(1) dan sukses(1) pada survey II pra *quick count*, maka peluang suksesnya dalam quick count sebesar 59.3% (Gambar 5).

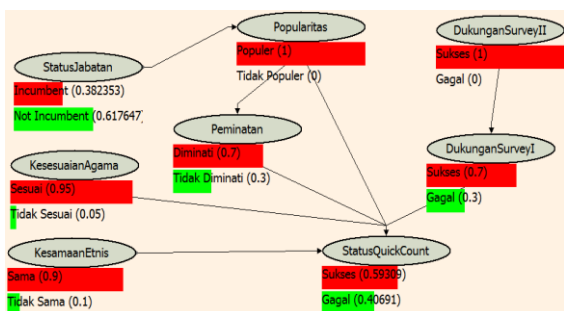


Figure 5. Evidence Popularitas Populer(1) dan DukunganSurveyII Sukses(1)

- Jika Status Jabatan *Incumbent*(1) dan sukses(1) pada survey II pra *quick count*, maka peluang suksesnya dalam quick count sebesar 58.9 % (Gambar 6).

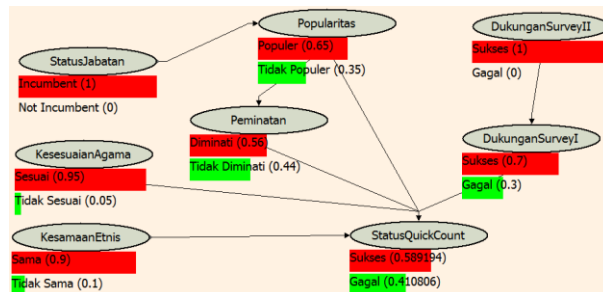


Figure 6. Evidence Incumbent(1) dan DukunganSurveyII Sukses(1)

Salah satu contoh seorang kandidat diprediksi tidak sukses dalam PILKADA dapat dilihat pada Gambar 7. Meskipun sebagian besar variable mempunyai nilai 1 (100%), namun kedua variable dukungan survey tidak diketahui, maka Status Quick Count Sukses hanya mendapatkan 44%.

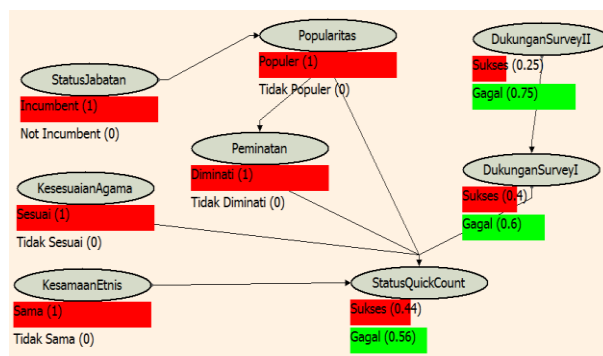


Figure 7. Dukungan Survey II dan I tidak mempunyai Evidence

KESIMPULAN

Penelitian ini merupakan simulasi untuk memprediksi pemenang PILKADA menggunakan *data mining* metode *Bayesian Network* dengan memanfaatkan *Tools Bayesian Network MSBNx*. Data diasumsikan merupakan hasil survey dari berbagai propensi, kabupaten/kota. Setelah data dikumpulkan, tahap selanjutnya adalah analisi data menggunakan *Bayesian network*. 4 tahap analisis data yaitu : menentukan variable dan kategori yang terkait dalam jaringan, membuat diagram

kausalitas, mengisi *conditional probability* setiap variable, dan memasukkan evidence untuk melihat hasil prediksi.

Kriteria minimum seorang kandidat untuk sukses dalam *quick count* yaitu menghasilkan peluang sukses diatas 50%, ditemukan bahwa variable Dukungan Survey I dan II sangat berpengaruh dalam pemenangan PILKADA, dimana jika kedua variable tersebut tidak mempunyai *evidence* atau tidak sukses, maka peluang *quick count* tidak akan sukses

DAFTAR PUSTAKA

- Kadie, Carl M., dkk, MSBNx: A Component-Cetric Toolkit for Modeling and Inference with Bayesian Network, Microsoft Research, Microsoft Cooperation One Micosoft Way, 28 July 2001, Redmond, WA 98052.
- Roselina, Yunita Nancy, dkk, Aplikasi Diagnosa Penyakit Asma Menggunakan Bayesian Network Berbasis Web, Jurnal Teknik Informatika, Vol. 1, September 2012, Riau.
- Santika, I Wayan, Pengembangan Sistem Pakar Konsultasi Hama dan Penyakit Tanaman Jeruk Menggunakan Metode Bayesian Network berbasis Web, Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI), vol. 1, No. 4, Agustus 2012
- Tama, Fitria Rani, Penerapan Bayesian Network untuk Memprediksi Pemenang Pilkada, Fakultas Matematika dan Ilmu Pengetahuan Alam ITB, 2008, Bogor.
- Variq, dkk, Sistem Tutor Cerdas Menggunakan Metode Bayesian Network, ITS-Undergraduate Paper, 2010, Surabaya.
- Yahdin, Sugandi, dkk, Aplikasi Pengambilan Keputusan Pada Perencanaan Produksi Berdasarkan Teorema Bayes, Media Informatika, No.1, Hal. 25-38, Juni 2008, Palembang.
- *) Penulis adalah Dosen Teknik Informatika STIMED Nusa Palapa Makassar**
- ***) Penulis adalah Dosen Dosen Teknik Informatika STMIK Handayani Makassar**