

**IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK
SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN
DAN TIDAK BERACUN**

SKRIPSI



MARSELIA GHANYYU WAHDINI

H13116516

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

2020

**IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK
SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN DAN
TIDAK BERACUN**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Ilmu Komputer Departemen Matematika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin**

Marselia Ghannyu Wahdini

H13116516

**PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

NOVEMBER 2020

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK
SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN
DAN TIDAK BERACUN**

adalah benar hasil karya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 27 November 2020



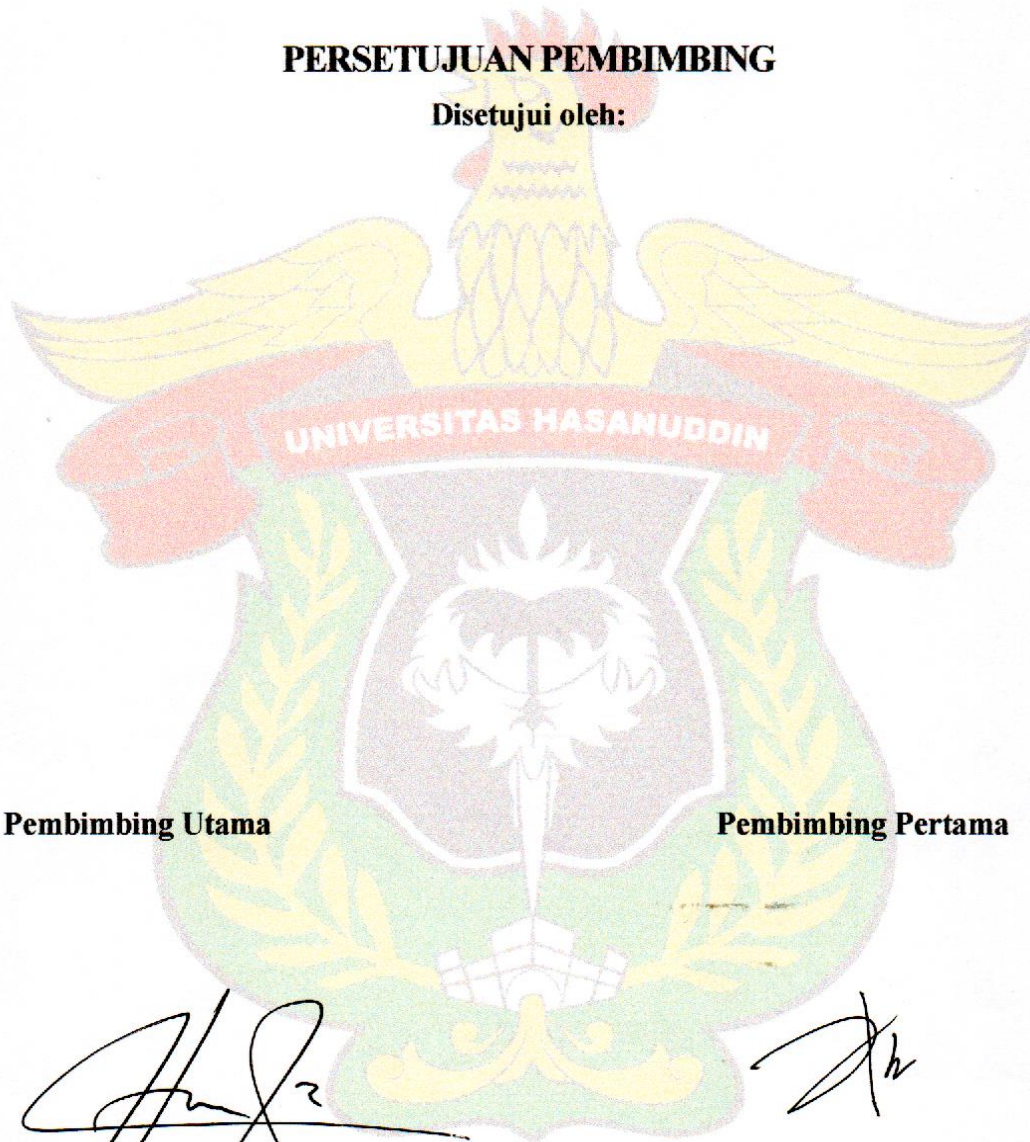
MARSELIA GHANYU WAHDINI

NIM. H13116516

**IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK
SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN
DAN TIDAK BERACUN**

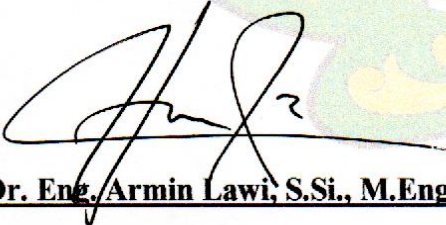
PERSETUJUAN PEMBIMBING

Disetujui oleh:




Pembimbing Utama

Pembimbing Pertama


Dr. Eng. Armin Lawi, S.Si., M.Eng.

NIP. 19720423 199512 1 001


Dr. Hendra, S.Si., M.Kom.

NIP. 19760102 200212 1 001

Pada 27 November 2020

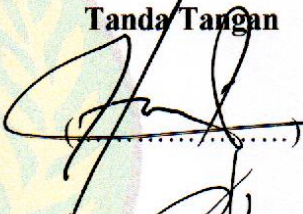
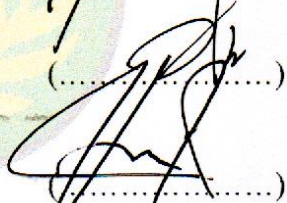
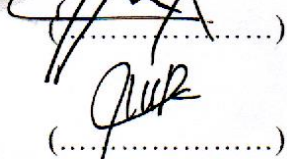
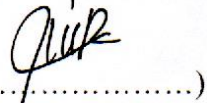
HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Marselia Ghanyyu Wahdini
NIM : H13116516
Program Studi : Ilmu Komputer
Judul Skripsi : Implementasi Metode Bayesian Network untuk Sistem
Rekomendasi Klasifikasi Jamur Beracun dan Tidak
Beracun

Telah berhasil mempertahankan di hadapan dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
1. Ketua	: Dr. Eng. Armin Lawi, S.Si., M.Eng.	
2. Sekretaris	: Dr. Hendra, S.Si., M.Kom.	
3. Anggota	: Dr. Muhammad Hasbi, M.Sc.	
4. Anggota	: Andi Muhammad Anwar, S.Si., M.Si.	

Ditetapkan di : Makassar

Tanggal : 27 November 2020



KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wa ta'ala*, Tuhan alam semesta yang telah memberikan nikmat kesempatan, kesehatan dan kemampuan sehingga penulisan skripsi ini bisa selesai. Shalawat serta salam senantiasa tercurah kepada *Rasulullah* Muhammad *Shallallahu Alaihi Wasallam*, yang merupakan teladan dalam menjalankan kehidupan di dunia.

Alhamdulillah, skripsi dengan Judul **“Implementasi Metode Bayesian Network untuk Sistem Rekomendasi Klasifikasi Jamur Beracun dan Tidak Beracun”** yang disusun sebagai salah satu syarat akademik untuk meraih gelar Sarjana Sains pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat diselesaikan. Tentunya, dalam penulisan skripsi ini, penulis mampu menyelesaikan tepat waktu berkat bantuan dan dukungan dari berbagai pihak. Oleh karena itu, ucapan terima kasih dan apresiasi yang tak terhingga kepada kedua orang tua penulis, Ayahanda **Agus Maryudhi** dan Ibunda **Yulia Asni Kurniawati** yang tak kenal lelah dalam memanjatkan doa serta memberikan nasihat dan motivasi kepada penulis. Tidak lupa pula terima kasih kepada kakak **Muhammad Maralifian Baasith**, dan adik **Maryam Arwa Balqista** yang selalu ada disaat senang maupun duka. Tugas akhir ini hanya setitik kebahagiaan kecil yang bisa penulis persembahkan.

Terima kasih juga penulis ucapkan kepada:

1. Rektor Universitas Hasanuddin, Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu** beserta jajarannya.
2. Dekan Fakultas Matematika dan Ilmu Pengetahuna Alam (FMIPA), **Dr.Eng. Amiruddin** beserta jajarannya.
3. Bapak **Dr. Nurdin, S.Si., M.Si.**, sebagai Ketua Departemen Matematika FMIPA Unhas. Penulis berterima kasih atas dedikasi dosen-dosen pengajar, serta Dapartemen atas ilmu dan bantuan yang bermanfaat.
4. Bapak **Dr. Muhammad Hasbi, M.Sc.**, sebagai Ketua Program Studi Ilmu Komputer FMIPA Unhas Penulis juga berterima kasih atas dedikasi dosen-dosen pengajar atas ilmu dan bantuan yang bermanfaat.
5. Bapak. **Dr. Eng, Armin Lawi, S.Si., M.Eng.**, sebagai dosen pembimbing utama sekaligus ketua tim penguji atas semua ilmu yang telah diberikan selama

proses perkuliahan dan senantiasa memotivasi penulis dalam penulisan skripsi ini.

6. Bapak **Dr. Hendra, S.Si.,M.Kom**, sebagai dosen pembimbing pertama sekaligus sekteraris tim penguji atas ilmu yang diberikan selama proses perkuliahan dan bimbingan, serta segala bentuk bantuan yang telah diberikan dalam penyusunan skripsi ini.
7. Bapak **Dr. Muhammad Hasbi, M.Sc.**, sebagai anggota tim penguji atas segala ilmu yang telah diberikan selama proses perkuliahan serta berbagai masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
8. Bapak **Andi Muhammad Anwar, S.Si., M.Si.**, sebagai anggota tim penguji atas segala ilmu yang telah diberikan selama proses perkuliahan serta berbagai masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
9. Saudaraku **Zinedine Kahlil Gibran Zidane, S.Si.**, yang telah banyak membantu, memotivasi dan meluangkan waktunya kepada penulis dalam hal perkuliahan sehingga penulis dapat melewati masa-masa sulit selama perkuliahan.
10. Saudara-saudara ku Saudara **Ilmu Komputer 2016**, terkhusus kepada **Mutawally Sya'rawy, S.Si., Rio Mukhtarom, S.Si., dan Marfiadhi Putra, S.Si.**, serta teman-teman yang tidak dapat disebutkan satu persatu, yang telah sama-sama berjuang serta banyak memberikan kenangan, dan dukungan tanpa henti, dalam menyelesaikan skripsi ini.
11. Saudara **Bebas Squad, Berlian Adriani Putri, Suci Rahmadana Anwar, Rizka Syahfitri, Nirwana Sari Hamka, Nurmayulina, Ainun Mardiyah Istiqamah, S.Si, Nisrina Syadza Dewanty, Tasnia Akil, , dan St. Hestiana Kadir, S.Si** yang senantiasa memberikan banyak dukungan serta motivasi kepada penulis, selalu ada baik dalam suka, maupun duka yang dialami oleh penulis, serta selalu ada untuk memberikan solusi dalam permasalahan yang dialami oleh penulis.
12. Keluarga besar **KKN Jepang 2019** yang secara ikhlas dan tulus mengabdikan kepada masyarakat.

13. Keluarga besar **A16ORITMA 2016** atas segala bentuk dukungan dan bantuan selama proses perkuliahan. Semoga kesuksesan selalu kita dapatkan dalam setiap langkah-langkah kita.
14. Seluruh pihak yang tidak dapat disebutkan satu per satu atas segala bentuk kontribusi, partisipasi, serta motivasi yang diberikan kepada penulis selama ini. Semoga apa yang kita berikan, dilipatgandakan oleh Allah Subhana Wata'ala.

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga karya ilmiah ini memberikan manfaat untuk pembaca.

Makassar, 27 November 2020

Marselia Ghanyyu Wahdini

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Marselia Ghanyyu Wahdini
NIM : H13116516
Programa Studi : Ilmu Komputer
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul:

“IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN DAN TIDAK BERACUN”

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada 27 November 2020

Yang menyatakan

(Marselia Ghanyyu Wahdini)

ABSTRAK

Salah satu keanekaragaman hayati tersebut adalah jamur. Tidak semua jenis jamur dapat dimakan (*edible*), banyak pula jenis jamur yang beracun (*poisonous*). *Family Agaricus* dan *Lepiota* yang secara liar hidup di alam terbuka dengan bentuk yang beraneka ragam, warna yang bermacam-macam, serta sifat yang belum banyak diketahui pada umumnya bersifat racun. Untuk membedakan jamur *family Agaricus* dan *Lepiota* tidak beracun dan yang beracun didasarkan pada bentuk, sifat, dan keadaannya sangat sukar dilakukan. Hal ini dikarenakan adanya bentuk hampir sama dari spesies jamur *family Agaricus* dan *Lepiota* yang dapat di konsumsi dengan spesies jamur lain yang beracun. Penelitian ini bertujuan untuk mengklasifikasikan jamur *family Agaricus* dan *Lepiota* ke dalam kelas tidak beracun dan beracun. pada penelitian ini penulis menggunakan dataset *Mushroom Classification* yang bersifat kategorik hanya dapat diklasifikasikan oleh pengklasifikasi yang mampu memproses variabel kategorik, seperti: *Bayesian Network*. Model yang dihasilkan dari data *training* memiliki akurasi sebesar 99% dan hasil evaluasi dari data *testing* menghasilkan akurasi sebesar 98%. Dari hasil akurasi tersebut dapat dinyatakan bahwa model *Bayesian Network* dapat mengklasifikasi jamur beracun dan tidak beracun dengan sangat baik.

Kata kunci: *Bayesian Network*, jamur, klasifikasi, kemungkinan, peluang, beracun, tidak beracun.

ABSTRACT

One such biodiversity is mushrooms. Not all types of mushrooms edible many of them poisonous. The agaricus and lepiota families live in the wild with various shape, various color, and characteristics that are not widely known in general they are poisonous. to distinguish mushrooms the agaricus and lepiota families that edible and poisonous based of shape, nature, and condition it is very difficult to do. This is because there is almost the same form of the species of fungi in the Agaricus and Lepiota families edible with other poisonous mushroom species. This is study aims to classify the Agaricus and Lepiota families fungi into edible and poisonous classes. The classification dataset of mushrooms that is categorical can only be classified by classifiers who are able to process categorical variable as Bayesian network. Models generated from training data has an accuracy of 99% and the result of the evaluations of the testing data has an accuracy of 98%. From the results of this accuracy, it can be stated that the Bayesian network model can classify poisonous mushrooms and mushrooms edible very well.

Keywords: Bayesian Network, Mushrooms, classification, opportunity, probability, poisonous and edible

DAFTAR ISI

IMPLEMENTASI METODE BAYESIAN NETWORK UNTUK SISTEM REKOMENDASI KLASIFIKASI JAMUR BERACUN DAN TIDAK BERACUN	ii
PERSETUJUAN PEMBIMBING.....	iv
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR	vi
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ix
ABSTRAK.....	x
ABSTRACT.....	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	1
DAFTAR TABEL.....	2
BAB I PENDAHULUAN.....	3
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Organisasi Skripsi.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Landasan Teori	6
2.1.1 Jamur	6
2.1.2 Teori Graf.....	10
2.1.3 Peluang.....	12
2.1.4 Teorema Bayes.....	14

2.1.5	Bayesian Network	15
2.1.6	Logaritma	17
2.1.7	Algoritma Chow-Liu	18
2.1.8	Confusion Matrix	24
2.2	State of the Art.....	25
2.3	Kerangka Konseptual	28
BAB III METODE PENELITIAN		29
3.1	Waktu dan Tempat	29
3.2	Deskripsi Data	29
3.3	Instrumen Penelitian.....	29
3.4	Tahapan Penelitian	30
3.4.1	Studi Literatur	30
3.4.2	Eksplorasi data	31
3.4.3	Preprocessing Data.....	31
3.4.4	Pembagian data	31
3.4.5	Training Data	31
3.4.6	Pembuatan model <i>Bayesian Network</i> menggunakan algoritma Chow-Liu	31
3.4.7	Evaluasi model.....	31
3.4.8	Performa dan Analisis Kinerja.....	31
3.4.9	Kesimpulan	31
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Eksplorasi Data.....	32
4.2	Preprocessing.....	40
4.2.1	Variabel Eleminasi	40
4.2.2	Label Encoding	41

4.3	Pembagian Data.....	41
4.4	Pembuatan Model <i>Bayesian Network</i> menggunakan algoritma Chow-Liu	41
4.4.1	Pembuatan graf.....	41
4.4.2	Pemberian arah pada graf.....	42
4.5	Variabel-Variabel yang Berpengaruh.....	43
4.6	Performa Dan Analisis Kinerja Model	46
4.7	Pembahasan	48
BAB V	KESIMPULAN DAN SARAN	50
5.1	Kesimpulan.....	50
5.2	Saran	50
	DAFTAR PUSTAKA	51
	LAMPIRAN	54
	Lampiran 1	54
	Lampiran 2	59
	Lampiran 3	67
	Lampiran 4	74
	Lampiran 5	82
	Lampiran 6	83

DAFTAR GAMBAR

Gambar 2.1 Struktur Jamur.....	6
Gambar 2.2 Klasifikasi Makhluk Hidup.....	7
Gambar 2.3 Jamur <i>Agaricus Hondesis</i>	8
Gambar 2.4 Jamur <i>Agaricus Bisporus</i>	8
Gambar 2.5 Jamur <i>Reddening Lepiota</i>	9
Gambar 2.6 Jamur <i>Lepiota Cristata</i>	9
Gambar 2.7 Graf.....	11
Gambar 2.8 <i>Acyclic Graph</i>	12
Gambar 2.9 <i>Directed Acyclic Graph</i>	12
Gambar 2.10 <i>Acyclic Graph</i> berdasarkan mutual informasi.....	21
Gambar 2.11 model <i>Bayesian Network</i> pada dataset contoh.....	24
Gambar 3.1 Diagram Alur Penelitian.....	30
Gambar 4.1 Distribusi Data 9 Variabel Pertama.....	38
Gambar 4.2 Distribusi Data Variabel 10 hingga Variabel 18	39
Gambar 4.3 Distribusi Data Variabel 19 Hingga Variabel 23	40
Gambar 4.4 Model <i>Bayesian Network</i> pada dataset asli.....	43
Gambar 4.5 Variabel-variabel berpengaruh pada model <i>Bayesian Network</i>	44

DAFTAR TABEL

Tabel 2.1 contoh dataset.....	19
Tabel 2.2 Mutual Informasi.....	20
Tabel 2.3 <i>log probability</i> kemungkinan 1	22
Tabel 2.4 Kemungkinan arah pada graf	23
Tabel 2.5 <i>Confusion Matrix</i>	24
Tabel 4.1 Variabel Data	32
Tabel 4.3 Variabel dan anggota variabel pada suatu sampel pada data training.....	44
Tabel 4.4 <i>Confusion Martix Training</i>	46
Tabel 4.5 <i>Training</i> data.....	47
Tabel 4.6 <i>Confusion Matrix Testing</i>	47
Tabel 4.7 <i>Testing</i> data	48

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara tropis yang memiliki sumber daya alam yang kaya akan keanekaragaman hayati. Salah satu keanekaragaman hayati tersebut adalah jamur. (Annisa, Ekamawanti, & Wahdina, 2017). Diperkirakan terdapat 1,5 juta spesies jamur di dunia, jenis yang teridentifikasi sebanyak 28.700 jenis jamur makroskopis (memiliki tubuh buah), jamur mikroskopis (tidak memiliki tubuh buah) sebanyak 24.000 dan 13.500 jenis lumut kerak (asosiasi simbiotik antara fungi dan alga), sedangkan jenis jamur yang belum teridentifikasi sejumlah 1.433.800 jenis, baik makro maupun mikro (Thomas, 2002).

Seiring dengan berkembangnya waktu, telah diketahui bahwa terdapat ribuan jamur dengan berbagai jenis. Tidak semua jenis jamur tidak beracun (*edible*). Banyak pula jenis jamur yang beracun (*poisonous*). *Family Agaricus* dan *Lepiota* yang secara liar hidup di alam terbuka dengan bentuk yang beraneka ragam, warna yang bermacam-macam, serta sifat yang belum banyak diketahui, pada umumnya bersifat racun. *Family Agaricus* dan *Lepiota* yang beracun dapat menyebabkan sakit pada seseorang yang memakannya sehingga dapat menyebabkan kematian. *Family Agaricus* dan *Lepiota* yang hidup di alam liar ada juga tidak beracun bahkan dijadikan sebagai obat (Putra, 2008).

Untuk membedakan jamur *family Agaricus* dan *Lepiota* tidak beracun dan yang beracun didasarkan pada bentuk, sifat, dan keadaannya, sangat sukar dilakukan. Hal ini dikarenakan adanya bentuk hampir sama dari spesies jamur *family Agaricus* dan *Lepiota* yang dapat di konsumsi dengan spesies jamur lain yang beracun. Penelitian ini bertujuan untuk mengklasifikasikan jamur *family Agaricus* dan *Lepiota* ke dalam kelas tidak beracun dan beracun (Putra, 2008). Menggunakan dataset *Mushroom Classification* yang bersifat kategorik hanya dapat diklasifikasikan oleh pengklasifikasi yang mampu memproses variabel kategorik, seperti *naïve Bayes*, *Bayesian Network*,

voting feature interval, *artificial neural network*, dan *adaptive neural fuzzy inference system*.

Beberapa penelitian telah dilakukan dengan berbagai metode berbeda, yaitu: *naïve Bayes* memiliki akurasi sebesar 99,85%, *voting feature interval* memiliki akurasi sebesar 84,53% (Wibowo, Rahayu, Riyanto, & Hidayatulloh, 2018), *artificial neural network* memiliki akurasi sebesar 70%, *adaptive neural fuzzy inference* memiliki akurasi 80% (Verma & Dutta, 2018).

Dalam penelitian ini penulis menggunakan metode *Bayesian Network*. Metode *Bayesian Network* dapat menangani variabel input yang saling berelasi dan dapat melakukan perhitungan peluang dari pengenalan jamur *family Agaricus* dan *Lepiota* berdasarkan ciri-cirinya.

Berdasarkan uraian di atas, penulis ingin melakukan penelitian mengenai **“Implementasi Metode Bayesian Network untuk Sistem Rekomendasi Klasifikasi Jamur Beracun dan Tidak Beracun”**

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas dapat dirumuskan:

1. Bagaimana membuat model *Bayesian Network* menggunakan algoritma Chow-Liu untuk klasifikasi data jamur berdasarkan ciri-cirinya?
2. Bagaimana menentukan variabel berpengaruh pada klasifikasi data jamur?
3. Bagaimana akurasi dari model *Bayesian Network* untuk kasifikasi data jamur?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Merupakan masalah klasifikasi dengan pembelajaran terawasi (*supervised learning*).
2. Dataset yang digunakan adalah dataset *Mushroom Classification family Agaricus and Lepiota*
3. Variabel-variabel pada dataset merupakan tipe kategorik dengan kelas biner.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah:

1. Membuat model *Bayesian Network* yang dapat mengklasifikasi jamur yang beracun dan tidak beracun
2. Dapat mengetahui variabel-variabel yang berpengaruh pada klasifikasi jamur beracun dan jamur tidak beracun.
3. Mengetahui akurasi dari model *Bayesian Network* pada dataset *Mushroom Classification*.

1.5 Organisasi Skripsi

Seiring berkembangnya waktu, telah diketahui bahwa terdapat ribuan jamur dengan berbagai jenis. Untuk membedakan jamur beracun dan tidak beracun didasarkan pada bentuk, sifat, dan keadaanya sangat sukar dilakukan. Hal ini dikarenakan adanya bentuk hampir sama dari spesies jamur *family Agaricus* dan *Lepiota* yang dapat di konsumsi dengan spesies jamur lain yang beracun.

Tujuan dari penelitian ini untuk mengklasifikasi jamur beracun dan tidak beracun berdasarkan ciri-ciri menggunakan metode *Bayesian Network* yang merupakan salah satu *Probabilistic Graphical Model* sederhana dengan struktur graf disebut dengan *Directed Acyclic Graph* dibuat menggunakan algoritma Chow-Liu dengan menghitung nilai mutual informasi dari setiap pasangan variabel acak kemudian graf yang telah dibuat akan diberikan arah secara acak sehingga menghasilkan *Directed Acyclic Graph* lebih dari satu, dalam penentuan model *Bayesian Network* ialah memilih satu dari sekian banyak *Directed Acyclic Graph* yang didapatkan dari graf yang sudah dibuat sebelumnya dengan menghitung *log probability* yang paling tinggi diantara nilai *log probability* dari setiap *Directed Acyclic Graph*. Dalam pembuatan model dibutuhkan 80% data *training* untuk membuat model dan 20% data *testing* digunakan untuk menguji seberapa baik model yang dibuat sehingga dapat mengetahui variabel-variabel apa saja yang berpengaruh dan mengetahui nilai akurasi dari model tersebut.

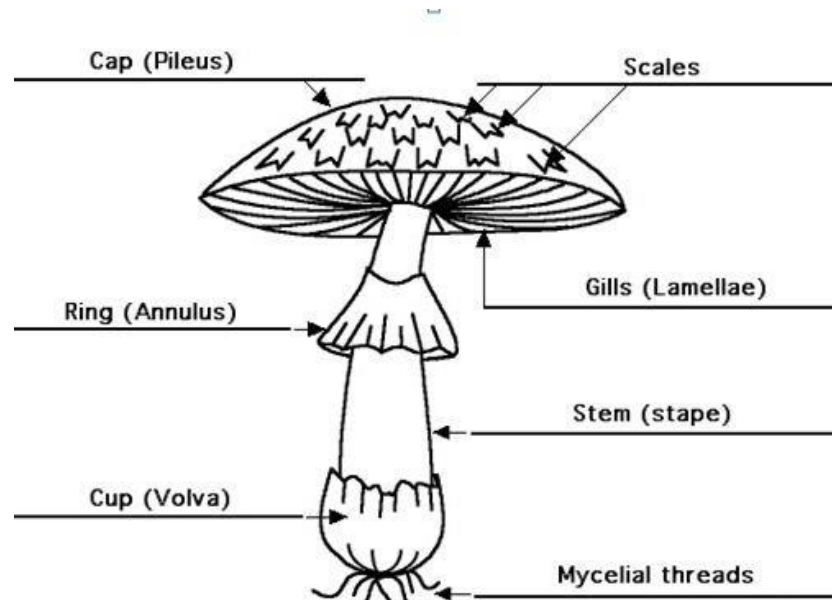
BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Jamur

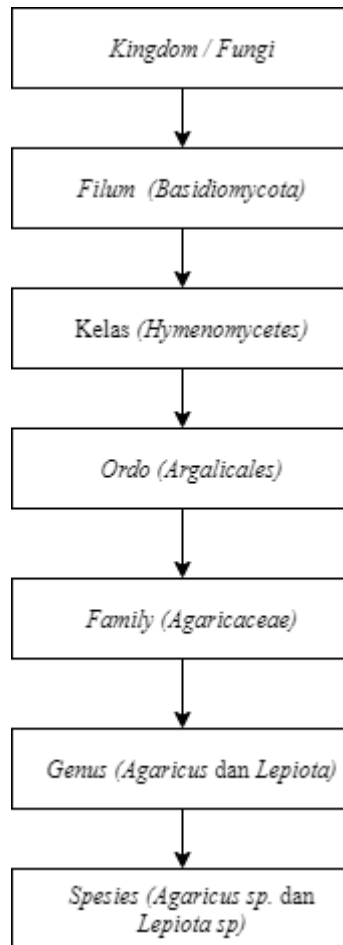
Jamur adalah tubuh-buah yang tampak di permukaan media tumbuh dari sekelompok *fungi* (*Basidiomycota*) yang berbentuk seperti payung, terdiri dari bagian yang tegak (batang) dan bagian yang mendatar atau membulat (Ulloa & Richard, 2000). Jamur umumnya menempati berbagai tipe habitat yaitu tanah, kayu, serasah, kotoran hewan dan sebagainya. Tipe ekosistem yang dapat ditumbuhi jamur ialah hutan, karena hutan memiliki tingkat kelembapan yang tinggi sehingga jamur mudah beradaptasi. Struktur jamur dapat dilihat pada gambar 2.1.



Gambar 2.1 Struktur Jamur

Dalam klasifikasi tumbuhan, *kingdom/fungi* dibagi kedalam empat *filum* dari *kingdom fungi* yaitu: *Chytridiomycota*, *Ascomycota*, *Zygomycota*, dan *Basidiomycota* (Ulloa & Richard, 2000). Setiap *filum* dibagi ke dalam kelas, setiap kelas dibagi menjadi ke dalam *ordo* dan setiap *ordo* dibagi menjadi *family*. Contohnya: jamur yang dibudidayakan (*Agaricus bisporus*) termasuk ke dalam *filum Basidiomycota*, *Agaricus*

bisporus termasuk ke dalam kelas *Hymenomycetes*. Kelompok dari kelas ini juga didasarkan pada karakteristik spora yang mikroskopik dan cara pelepasan spora. *Agaricales* merupakan ordo dari *Agaricus bisporus* (terdiri semua jamur yang memiliki insang) dan jamur ini termasuk ke dalam *family Agaricaceae* dimana merupakan *spesies Agaricus*, selanjutnya dibagi ke dalam *genus (singular, genus)* dan *genus* dibagi menjadi ke dalam *spesies*, klasifikasi makhluk hidup dapat dilihat pada gambar 2.2 (Lincoff, 1981).



Gambar 2.2 Klasifikasi Makhluk Hidup

2.1.1.1 Family Agaricus dan Lepiota (Agaricaceae)

Agaricaceae merupakan *family* dari jamur yang umumnya dibudidayakan (*Agaricus bisporus*) dan biasanya banyak ditemukan pada daerah kota dan pinggiran kota.

Walaupun demikian *Agaricaceae* banyak tumbuh di hutan, batang kayu, daerah yang ditutupi rumput (*Grass area*), halaman rumput dan disepanjang pinggir jalan. Sebagian dari *Agaricaceae* merupakan jamur yang baik untuk dikonsumsi dan banyak dibudidayakan ialah *Agaricus bisporus*, dengan ciri-ciri yang berbentuk hampir bulat seperti kancing, berwarna putih bersih, krem, atau coklat muda, sebagian lain adalah jamur yang beracun dan sebagian kecil jamur yang mematikan. Delapan *genus* ditemukan di *North America*: *Agaricus* (\pm 200 spesies), *Chlorophyllum* (1 spesies), *Cystoderma* (20 spesies), *Dissoderma* (1 spesies), *Melanophyllum* (1 spesies), *Lepiota* (\pm 100 spesies), *Phaeolepiota* (1 spesies), dan *Aquamanita* (2 spesies). Ciri-ciri *Agaricaceae*, yaitu: memiliki sisik pada butiran-butiran kecil di daerah tudung dan insang yang terpisah dari batang. *Agaricaceae* semuanya memiliki tudung *membrane* dan kebanyakan diantaranya memiliki cincin yang melekat pada batang (Putra, 2008).

Salah satu jenis jamur dari *genus Agaricus* yang beracun dapat dilihat pada gambar 2.3 dan jamur tidak beracun tertera pada gambar 2.4.



Gambar 2.3 Jamur *Agaricus Hondesis*



Gambar 2.4 Jamur *Agaricus Bisporus*

Pada genus seperti *Lepiota* dan *Melanophyllum*, cincin sering tidak ditemukan (hilang), cetakan spora kemungkinan berwarna putih, hijau gelap, coklat tua, kemerah-merahan, keungu-unguan, dan coklat. Jamur ini dikelompokkan dalam *single family* atas dasar sebuah hubungan dari karakteristik bentuk mikroskopik dan bahan kimia, seperti daging insang yang tidak pernah berlainan (seperti *amanita*), spora biasanya halus, berdinding tipis, dan dibagian ujungnya memiliki pori-pori (Lincoff, 1981).

Salah satu jenis jamur dari genus *Lepiota* yang bisa dikonsumsi seperti pada gambar 2.5 dan yang beracun seperti pada gambar 2.6.



Gambar 2.5 Jamur *Reddening Lepiota*



Gambar 2.6 Jamur *Lepiota Cristata*

Beberapa jenis *family Agaricus* dan *Lepiota* termasuk dalam makro *fungi* karena memiliki bentuk besar, dapat dilihat dengan mata telanjang dan dapat dipegang dengan tangan (Ulloa & Richard, 2000). Salah satu jamur beracun dari jenis *lepiota cristata*,

memiliki ciri-ciri mempunyai tubuh buah berbentuk payung, bertekstur mudah hancur, *lamella* berbentuk insang dengan warna putih, warna tubuh buah *orange* kemerahan dengan bagian tengah berwarna lebih gelap, tidak memiliki cincin dan cawan (Mahendra, 2017).

2.1.1.2 Ciri-ciri Jamur Beracun

Untuk menentukan jamur kedalam kelas tidak beracun atau beracun sangat sukar dilakukan, adapun beberapa ciri-ciri Jamur beracun

1. Jenis jamur beracun pada umumnya mempunyai warna yang mencolok merah-darah, hitam-legam, biru-tua, ataupun warna-warna lainnya.
2. Jenis jamur beracun dapat menghasilkan bau yang menusuk hidung, seperti bau telur busuk.
3. Jenis jamur beracun mempunyai cincin atau cawan walaupun ada yang sebaliknya.
4. Jenis jamur beracun umumnya tumbuh pada tempat yang kotor.
5. Jenis jamur beracun cepat sekali berubah warna, misal dari putih ke warna gelap kalau dimasak ataupun dipanaskan.
6. Kalau jenis jamur beracun dikerat oleh pisau yang terbuat dari perak, atau dikerat oleh pisau biasa kemudian benda perak didekatkan dengan keratan tadi, maka benda perak terbentuk warna hitam atau biru, itu menandakan bahwa jamur tersebut beracun (Hendritomo & Istawan, 2010).

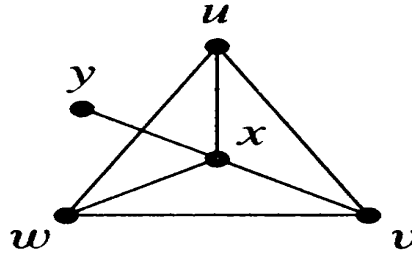
2.1.2 Teori Graf

Graf G adalah pasangan himpunan (V,E) dimana V adalah himpunan tak kosong dan berhingga dari obyek-obyek yang disebut **simpul** dan E adalah himpunan pasangan tak berurut (bisa kosong) dari elemen berlainan dari V yang disebut **sisi**. Istilah lain untuk simpul adalah titik atau *vertex* atau *node*, sedangkan sisi biasa juga disebut busur atau garis atau *edge*. Sebuah graf direpresentasikan dalam sebuah gambar/diagram dimana simpul dilambangkan dengan noktah, lingkaran atau titik tebal yang ditandai dengan angka atau huruf, sedangkan sisi dilambangkan dengan ruas garis atau kurva (berarah atau tidak berarah) yang menghubungkan pasangan simpul.

Sebuah graf dengan himpunan simpul V dan himpunan sisi E yang diberikan, dapat dilihat pada gambar 2.7, sebagai berikut

$$V = \{u, v, w, x, y\}$$

$$E = \{uv, uw, ux, vw, vx, xw, xy\}$$



Gambar 2.7 Graf

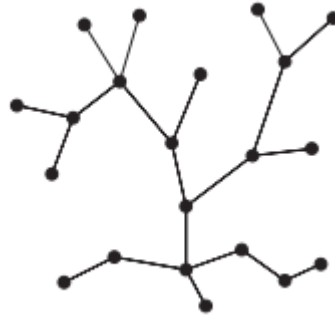
Himpunan simpul dari G dinotasikan dengan $V(G)$, sedangkan himpunan sisi dari G dinotasikan dengan $E(G)$ (Harris, Hirst, & Mossinghoff, 2008).

2.1.2.1 Directed Graph (*Graf Berarah*)

Graf berarah adalah graf dengan *edge* yang memiliki satu arah antar *node* yang dihubungkannya. Pada graf berarah sisi lebih sering disebut busur. Busur (V_j, V_k) berbeda dengan busur (V_k, V_j) . Jika (V_j, V_k) adalah sisi berarah/busur, maka V_j adalah simpul asal dan V_k adalah simpul terminal. Pada busur (V_j, V_k) terdapat jalur dari V_j ke V_k , tapi tidak sebaliknya (Waisakurnia, 2013).

2.1.2.2 Acyclic Graph

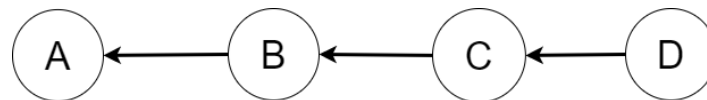
Acyclic graph adalah graf yang tidak memuat siklus. *Acyclic graph* dapat menghubungkan sebuah simpul lama dengan sebuah simpul baru dengan syarat tidak menghasilkan lingkaran. Untuk lebih jelasnya, dapat dilihat pada gambar 2.8 (Diestel, 1997).



Gambar 2.8 *Acyclic Graph*

2.1.2.3 Directed Acyclic Graph

Directed Acyclic Graph adalah graf yang memiliki arah (graf berarah) dan tidak memiliki siklus didalamnya, (Waisakurnia, 2013). Dalam hal ini siklus yang dimaksud ialah siklus tak berarah, dapat dilihat pada gambar 2.9



Gambar 2.9 *Directed Acyclic Graph*

2.1.3 Peluang

Kemungkinan terjadinya suatu kejadian sebagai hasil percobaan statistika dinilai dengan menggunakan himpunan bilangan real disebut **bobot** atau **peluang** bernilai dari 0 sampai 1. Untuk setiap titik pada ruang sampel dikaitkan suatu peluang sedemikian rupa sehingga jumlah semua bobot sama dengan 1. Jika suatu titik sampel tertentu kemungkinan besar akan terjadi, maka bobotnya mendekati angka 1. Sebaliknya, bobot yang mendekati angka 0 diberikan pada titik sampel kemungkinan kecil akan terjadi (Walpole & Myers, 1995). Dapat dirumuskan pada persamaan (1):

$$P = \left(\frac{n(A)}{n(S)} \right) \quad (1)$$

Dimana: P = Peluang

$n(A)$ = Jumlah kasus yang memenuhi syarat (ketentuan yang ditanyakan)

$n(S)$ = Jumlah total semua kasus

Contoh: terdapat 1 buah dadu berisis 6 kemudian dilempar 1 kali, berapakah peluang munculnya angka “5”

Diketahui:

$$n(S) = \{1,2,3,4,5,6\}$$

$$n(A) = \{5\}$$

Jadi: $P = \left(\frac{n(A)}{n(S)}\right) = \frac{1}{6}$, peluang angka “5” muncul pada sebuah dadu, ialah: $\frac{1}{6}$

2.1.3.1 Variabel Acak

Variabel acak adalah variabel yang nilainya bergantung dari hasil fenomena acak (Blitzstein & Hwang, 2019). Biasanya dinyatakan dengan huruf besar, misalnya X sedangkan nilainya akan dinyatakan dengan huruf kecil, misalnya x (Walpole & Myers, 1995).

Ditulis, pada persamaan (1)

$$f(X) = P(X = x) \tag{2}$$

Dimana fungsi $f(x)$ dinamakan fungsi peluang atau distribusi peluang.

2.1.3.2 Peluang Bersyarat

Peluang terjadinya suatu kejadian A bila diketahui bahwa kejadian B telah terjadi disebut peluang bersyarat dan dinyatakan dengan $P(B|A)$. Lambang $P(B|A)$ biasanya dibaca ‘peluang B terjadi bila A terjadi’ atau lebih sederhananya lagi ‘peluang B , bila A diketahui’. Peluang bersyarat B bila A diketahui dinyatakan dengan $P(B|A)$, ditentukan oleh persamaan (3) (Walpole & Myers, 1995).

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{3}$$

Dimana $P(B|A)$ = Peluang B bila A diketahui.

$P(A \cap B)$ = peluang A irisan B .

$$P(A) = \text{Peluang } A$$

Contoh: sebuah dadu bersisi 6 dilemparkan 1 kali dan diketahui mata dadu yang muncul adalah ganjil. Tentukan peluang akan muncul mata dadu yang lebih dari “4”

$$\text{Diketahui: } n(S) = \{1,2,3,4,5,6\}$$

$$P(A) = \text{ganjil } \{1,3,5\} = \frac{3}{6} = \frac{1}{2}$$

$$P(B) = \text{lebih dari “4” } \{5,6\} = \frac{2}{6} = \frac{1}{3}$$

$$P(A \cap B) = \text{ganjil dan lebih dari “4” } \{5\} = \frac{1}{6}$$

$$\text{Jadi: } P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{6} \times \frac{2}{1} = \frac{1}{3}$$

2.1.4 Teorema Bayes

Teorema Bayes digunakan untuk menghitung peluang terjadinya suatu peristiwa berdasarkan pengaruh yang diperoleh dari hasil observasi. Disamping ini, metode Bayes memanfaatkan data sampel yang diperoleh dari populasi juga memperhitungkan suatu distribusi awal yang disebut *prior*. Setelah pengamatan dilakukan, informasi dalam distribusi *prior* dikombinasikan dengan data sampel melalui teorema Bayes. Bila seseorang mengamati kejadian dan mempunyai keyakinan bahwa ada kemungkinan B akan muncul maka peluang disebut *prior*, sedangkan ada informasi tambahan bahwa misalkan kejadian A telah muncul mungkin akan terjadi perubahan terhadap perkiraan semula mengenai kemungkinan B untuk muncul. Peluang untuk B sekarang adalah peluang bersyarat akibat A dan disebut sebagai peluang *posterior*. Teorema Bayes merupakan mekanisme untuk memperbaharui peluang *prior* menjadi peluang *posterior*. Thomas Bayes menggambarkan hubungan antara peluang bersyarat dari dua kejadian merupakan salah satu cara yang baik mengatasi ketidakpastian data dengan menggunakan formula Bayes yang dinyatakan dengan rumus pada persamaan (4) sebagai berikut (Saputra, Dahria, & Putri, 2017).

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (4)$$

Dimana:

A dan B = kejadian

$P(A|B)$ = peluang A diberikan B benar

$P(B|A)$ = peluang B diberikan A benar

$P(A)$ dan $P(B)$ = peluang A dan B

2.1.5 Bayesian Network

Bayesian Network merupakan salah satu *Probabilistic Graphical Model* (PGM) sederhana yang dibangun dari teori peluang dan teori graf. Teori *probabilistic* berhubungan langsung dengan data, sedangkan teori graf berhubungan langsung dengan representasi yang ingin diperoleh (Heckerman, 1986).

Metode *Bayesian Network* merupakan metode yang baik di dalam *mechine learning* berdasarkan data *training* dengan menggunakan peluang bersyarat sebagai dasarnya. *Bayesian Network* terdiri dari dua bagian utama, yaitu

1. Struktur graf *Bayesian Network* disebut dengan *Directed Acyclic Graph* (DAG), 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) antara variabel yang dihubungkannya. Tidak adanya *edge* menandakan adanya hubungan kebebasan kondisional di antara variabel.
2. Pada *Bayesian Network* *node* berkorespondensi dengan variabel acak. Setiap *node* diasosiasikan dengan himpunan peluang bersyarat $P(X_i|X_{j(i)})$ sehingga X_i adalah variabel yang diasosiasikan dengan *node* dan $X_{j(i)}$ adalah parent dalam *graph*.

Misalkan himpunan dari *node* dinyatakan dengan (X_1, X_2, \dots, X_N) , jika terdapat *edge* dari *node* X_j ke *node* X_k dikatakan bahwa X_j adalah *parent* dari X_k , dan X_k adalah *child*

dari X_j . Berdasarkan gambar 2.9 *parent* X_4 adalah X_1 dan *child* untuk X_2 adalah X_1 . (Rachmat, 2008).

2.1.5.1 Inferensi Bayes

Inferensi Bayes adalah inferensi *statistical* dimana beberapa *evidence* atau pengamatan digunakan untuk menghitung peluang bahwa sebuah hipotesis itu benar.

Pada penggunaannya, inferensi Bayes dilakukan dengan cara melihat kesamaan pada hipotesis untuk menentukan hipotesis tersebut termasuk ke kelas mana dengan *evidence* yang ada. Hasilnya dapat diperoleh dengan melihat kesamaan hipotesis dengan kecocokan *evidence* yang ada dengan hipotesis. Inferensi Bayes dapat digunakan untuk membedakan dua hipotesis yang saling bertentangan (Aribowo, 2010).

2.1.5.2 Variabel Eliminasi

Diberikan model *Bayesian Network* X , dengan *evidence* E dan X_q adalah variabel yang ditanyakan, dinotasikan $P(X_q|E)$ biasanya hanya melibatkan sebagian kepadatan yang terkait dengan jaringan. Jika kepadatan $P(X_i|X_{j(i)})$ diperlukan untuk menjawab variabel yang dibutuhkan, maka $X_{m(i)}$ adalah *Requisite Variable*. Himpunan *Requisite Variable* dilambangkan dengan X_R . dapat dilihat pada persamaan (6).

$$P(X_q, E) = \sum_{\substack{X_R \\ X_q, X_E}} \left(\prod_{X_i \in X_R} P(X_i|X_{j(i)}) \right) \quad (6)$$

Dimana kepadatan peluang harus dibatasi untuk domain yang tidak mengandung *evidence*. Untuk jumlah dari *Requisite Variable* dilambangkan N yang tidak dalam variabel X_q . Misalkan terdapat variabel $\{X_1, X_2, X_3, \dots, X_n\}$ maka, dapat dilihat pada persamaan (7).

$$P(X_q, E) = \sum_{X_N} \dots \sum_{X_1} P(X_N|X_{j(N)}) \times \dots \times P(X_1|X_{j(1)}) \quad (7)$$

Karena, X_1 hanya terdapat pada $P(X_1|X_{j(1)})$. Maka persamaan (7) dapat disederhanakan, menjadi persamaan (8).

$$P(X_q, E) = \sum_{X_N} \dots \sum_{X_2} P(X_N|X_{j(N)}) \times \dots \times P(X_2|X_{j(2)}) \times \sum_{X_1} P(X_1|X_{j(N)}) \quad (8)$$

Kemudian $\sum_{X_1} P(X_1|X_{1(j)})$ dijadikan fungsi, dapat dilihat pada persamaan (9).

$$f_1(X_1, X_{1(j)}) = \sum_{X_1} P(X_1|X_{1(j)}) \quad (9)$$

Maka, persamaan persamaan 7 menjadi persamaan (10).

$$P(X_q, E) = \sum_{X_N} \dots \sum_{X_2} P(X_N|X_{j(N)}) \times \dots \times P(X_2|X_{j(2)}) \times f_1(X_1, X_{1(j)}) \quad (10)$$

Hal yang serupa dilakukan kepada variabel X_2, X_3, \dots, X_{N-1} (Cozman, 2000).

2.1.6 Logaritma

Logaritma adalah hasil kali kebalikan (*invers*) dari sebuah perpangkatan. Jika sebuah perpangkatan, operasi logaritma berlaku untuk nilai $a > 0$. Rumus dasar logaritma dapat dilihat pada persamaan 5

$$\text{jika, } a = e, \text{ maka } {}^e \log c = \ln c$$

Dimana:

a = basis atau bilangan pokok logaritma

b = bilangan yang dicari logaritma

c = hasil operasi logaritma

\ln = logaritma *natural*

e = bilangan *euler*

x = bilangan rill

2.1.7 Algoritma Chow-Liu

Misalkan $P(X)$ adalah peluang gabungan dari n variabel diskrit x_1, x_2, \dots, x_n , dan X melambangkan n vector (x_1, x_2, \dots, x_n) . Hasil kali $P(X)$ didefinisikan sebagai hasil kali dari beberapa distribusi komponen dengan orde yang lebih rendah sedemikian rupa sehingga hasil kali tersebut merupakan perluasan peluang dari hasil kali distribusi yang lebih rendah. Perluasan $P(X)$ di tulis sebagai persamaan (11).

$$P_t(X) = \prod_{i=1}^n P(x_i | x_{j(i)}), \quad 0 \leq j(i) < i \quad (11)$$

Pasangan yang terdiri dari $X = \{x_i | i = 1, 2, \dots, n\}$ dan pemetaan $j(i)$ dengan $0 \leq j(i) < i$ disebut *dependence tree* (Lewis II, 1959).

Misalkan $P(X)$ dan $P_a(X)$ adalah dua distribusi peluang dari n variabel diskrit $X = (x_1, x_2, \dots, x_n)$. Maka diketahui

$$I(P, P_a) = \sum_x P(X) \ln \frac{P(X)}{P_a(X)}, \quad I(P, P_a) \geq 0 \quad (12)$$

Pada persamaan (12) akan dijadikan kriteria untuk mengaproksimasi distribusi *ordo n* pada *dependence tree*.

Diberikan distribusi peluang $P(x_1, x_2, \dots, x_n)$, x_i menjadi bilangan diskrit, untuk menemukan distribusi dari *dependence tree* $P_\tau(x_1, x_2, \dots, x_n)$ seperti yang $I(P, P_\tau) \leq I(P, P_t)$ untuk semua $t \in T_n$ adalah himpunan semua *dependence tree* urutan pertama. Solusi τ disebut *dependende tree* urutan pertama yang optimal. Beberapa definisi yang untuk menjelaskan solusi terhadap masalah optimisasi, sebagai berikut

Mutual information $I(X_i, X_j)$ antara dua variabel X_i dan X_j terdapat pada persamaan (13).

$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i, X_j) \ln \left(\frac{P(X_i, X_j)}{P(X_i)P(X_j)} \right) \quad (13)$$

Dari definisi umum mutual informasi diketahui $I(X_i, X_j)$ tidak *negative*. Bobot maksimum *dependende tree*, dalam representasi grafis dari hubungan ketergantungan, untuk setiap *node dependende tree* ditetapkan bobot *node* $I(X_i, X_j)$ diberikan *dependende tree* t . Sebuah *dependende tree* dengan bobot maksimum adalah suatu *dependende tree* t sedemikian sehingga seluruh t' pada T_n , memenuhi persamaan (14).

$$\sum_{i=1}^n I(X_i, X_j) \geq \sum_{i=1}^n I(X_i, X_j) \quad (14)$$

Masalah maksimalisasi dari $\sum_{i=1}^n I(X_i, X_j)$ pada persamaan (14) dapat diselesaikan tanpa mempertimbangkan semua ekspansi yang mungkin terjadi. Solusi langsung adalah kemungkinan masalahnya menemukan *dependende tree* urutan pertama yang optimal di transformasikan menjadi memaksimalkan total bobot *node dependende tree* karena bobot *node* bersifat dapat dijumlahkan, maka bobot *dependence tree* dapat dibangun *node* demi *node* (Chow & Liu, 1968).

Sebagai contoh, jika diberikan suatu data dengan 4 variabel yaitu A,B,C,D dengan 8 sampel sebagai berikut:

Tabel 2.1 contoh dataset

No	A	B	C	D
1	a	b	a	p
2	b	b	b	e
3	a	b	b	p
4	a	a	a	p
5	b	a	a	p
6	a	b	b	e
7	a	b	b	e
8	b	a	a	p

Pada tabel 2.1 terdapat contoh dataset yang akan dibuatkan model *Bayesian network* menggunakan algoritma Chow-Liu, berikut perhitungan nilai mutual informasi dari salah satu kombinasi variabel acak:

$$\begin{aligned}
 I(A, B) &= \sum_{A, B} P(A, B) \ln \left(\frac{P(A, B)}{P(A)P(B)} \right) \\
 &= P(A = 'a', B = 'b') \ln \frac{P(A = 'a', B = 'b')}{P(A = 'a') \times P(B = 'b')} + \\
 &P(A = 'a', B = 'a') \ln \frac{P(A = 'a', B = 'a')}{P(A = 'a') \times P(B = 'a')} + \\
 &P(A = 'b', B = 'a') \ln \frac{P(A = 'b', B = 'a')}{P(A = 'b') \times P(B = 'a')} + \\
 &P(A = 'b', B = 'b') \ln \frac{P(A = 'b', B = 'b')}{P(A = 'b') \times P(B = 'b')} \\
 &= 0,125 \ln \frac{0,125}{0,625 \times 0,375} + 0,5 \ln \frac{0,5}{0,625 \times 0,625} + 0,25 \ln \frac{0,25}{0,375 \times 0,375} \\
 &+ 0,125 \ln \frac{0,125}{0,625 \times 0,375} = 0,110
 \end{aligned}$$

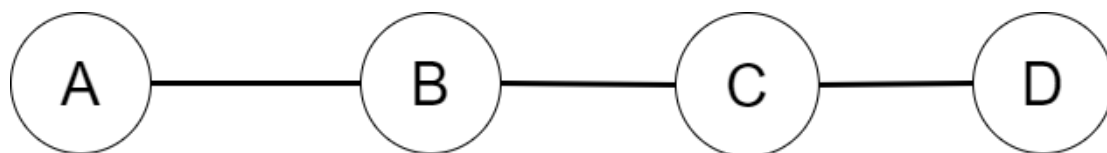
Prosedur diatas diulang untuk setiap kombinasi variabel acak dihitung sehingga menghasilkan nilai mutual informasi yang kemudian nilainya diurutkan dari nilai tertinggi hingga nilai terendah, dapat dilihat pada tabel 2.2

Tabel 2.2 Mutual Informasi

Mutual informasi	Nilai
$I(B, C)$	0,380
$I(C, D)$	0,380
$I(B, D)$	0,240
$I(A, B)$	0,110

Mutual informasi	Nilai
$I(A, C)$	0,033
$I(A, D)$	0,002

Karena, $I(B, C)$ dan $I(C, D)$ memiliki nilai yang besar, maka $I(B, C)$ dan $I(C, D)$ merupakan dua *edge* pertama dari *dependence tree*, untuk memilih *edge* berikutnya maka $I(B, D)$ ditambahkan *edge* maka menghasilkan siklus sehingga *edge* $I(B, D)$ tidak digunakan, untuk *edge* $I(A, B)$ ditambahkan *edge* tidak terjadi siklus maka *edge* tersebut digunakan, untuk $I(A, C)$ dan $I(A, D)$ jika ditambahkan *edge* terjadi siklus maka *edge* tersebut tidak digunakan, dapat dilihat pada gambar 2.9



Gambar 2.10 *Acyclic Graph* berdasarkan mutual informasi

Setelah didapatkan *Acyclic Graph* berdasarkan tabel 2.2 maka langkah selanjutnya ialah pemberian arah pada *Acyclic Graph* sehingga menghasilkan *Directed Acyclic Graph* untuk model *Bayesian Network*.

Jumlah kemungkinan *Directed Acyclic Graph* yang dapat dibentuk dari *acyclic graph* adalah $2^{\text{jumlah sisi}}$. Pada contoh ini jumlah kemungkinan adalah $2^3 = 8$ *Directed Acyclic Graph*. Di antara 8 kemungkinan *Directed Acyclic Graph* hanya 1 yang dapat dijadikan model *Bayesian Network* yang memiliki nilai *log probability* paling tinggi. Untuk menghitung nilai *log probability* pada setiap kemungkinan *Directed Acyclic Graph*, sebagai berikut:

Untuk sampel pertama pada dataset, nilai *log probability* dari kemungkinan 1 adalah

$$\log(P(\mathbf{X})) = \ln(P(D) \times P(C|D) \times P(B|C) \times P(A|B))$$

$$\begin{aligned} \ln(P(X)) &= \ln(P(D = 'p') \times P(C = 'a'|D = 'p') \times P(B = 'b'|C = 'a') \\ &\quad \times P(B = 'b'|A = 'a')) \\ &= \ln(P(X)) = \ln(0,625 \times 0,8 \times 0,25 \times 0,8) \\ &= \ln(P(X)) = \ln 0,1 \\ &= \ln(P(X)) = -2,30 \end{aligned}$$

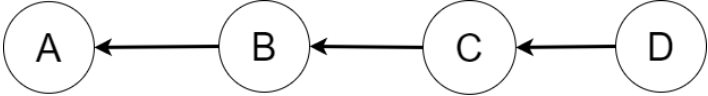
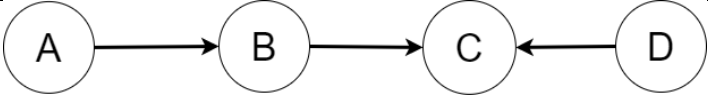
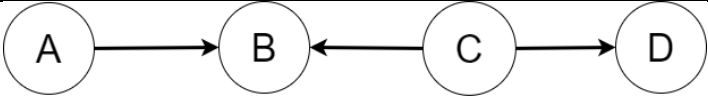
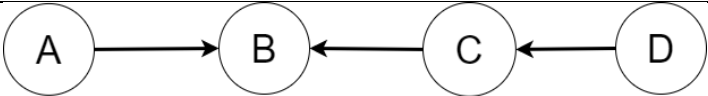
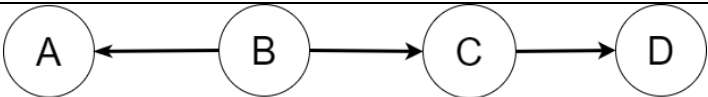
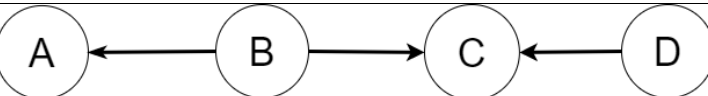
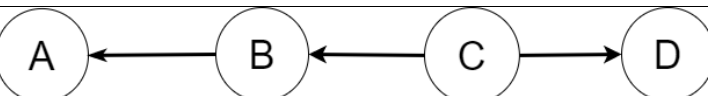
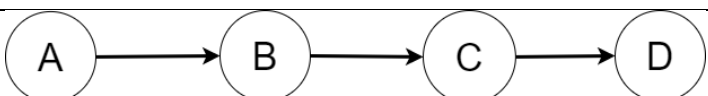
Prosedur yang sama dilakukan pada sampel kedua sampai sampel kedelapan, *Conditional Probability Table* (CPT), logaritma yang digunakan ialah logaritma *natural (euler)* dapat dilihat pada lampiran 5.

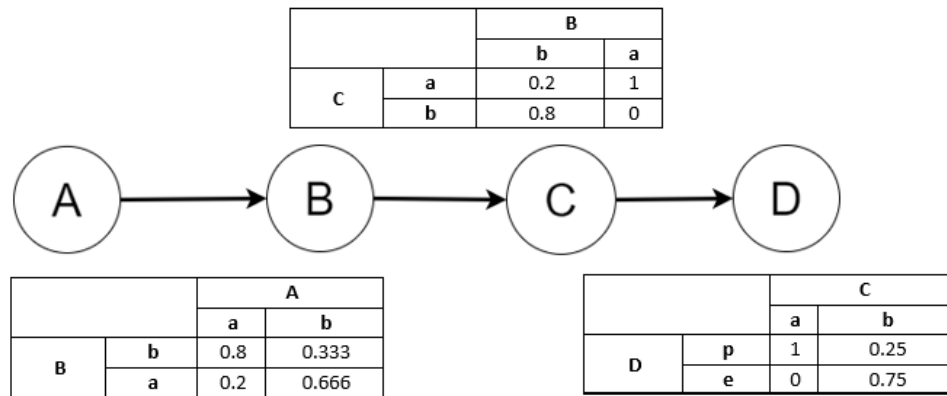
Tabel 2.3 *log probability* kemungkinan 1

No	$P(X)$	$\ln(P(X))$
1	0,10	-2,30
2	0,07	-2,56
3	0,10	-2,30
4	0,12	-2,12
5	0,25	-1,38
6	0,30	-1,20
7	0,30	-1,20
8	0,24	-1,42
Total		-7,48

Prosedur yang sama dilakukan pada kemungkinan kedua sampai kemungkinan delapan sehingga, menghasilkan tabel 2.4

Tabel 2.4 Kemungkinan arah pada graf

No	Kemungkinan arah	<i>Log probability</i>
1		-7,48
2		-20,00
3		-55,89
4		-6,53
5		-6,19
6		-6,69
7		-7,18
8		-4,44



Gambar 2.11 model *Bayesian Network* pada dataset contoh

2.1.8 Confusion Matrix

Di dalam *machine learning*, mengukur kinerja atau performa dari suatu model adalah hal yang esensial. Model yang diperoleh dari pelatihan melalui data *training* perlu diuji melalui data *testing*. Kinerja diukur berdasarkan seberapa baik model tersebut memprediksi dengan benar data yang ada.

Pada klasifikasi biner, kelas positif yang berhasil diprediksi dengan benar disebut *true positive*, jika kelas positif tersebut diprediksi negatif (salah) disebut *false negative*. Kelas negatif yang berhasil diprediksi negatif (benar) disebut *true negative* dan kelas negatif yang diprediksi positif disebut *false positive*. Jumlah dari kasus-kasus tersebut direpresentasikan dalam suatu tabel kontingensi yang disebut *confusion matrix* (Sweet, 1988). Dapat dilihat pada tabel 2.5.

Tabel 2.5 *Confusion Matrix*

		Kelas asli	
		Positif	Negatif
Hasil Prediksi	Positif	<i>True Positive</i>	<i>False Negative</i>
	Negatif	<i>False Positive</i>	<i>True Negative</i>

Dari tabel 2.2 menghasilkan akurasi, presisi, sensitivitas, dan *F1-Score*. Akurasi adalah ukuran kinerja yang menunjukkan seberapa baik suatu pengklasifikasi dalam mengklasifikasikan seluruh data. Akurasi adalah rasio antara observasi yang diklasifikasikan secara benar dengan total observasi, dapat dilihat pada persamaan (15).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (15)$$

Presisi adalah ukuran kinerja yang menunjukkan seberapa besar kebenaran suatu pengklasifikasi dari seluruh kelas positif yang diprediksi. Presisi adalah rasio antara jumlah kelas positif yang diklasifikasikan secara benar dengan jumlah observasi yang diklasifikasikan positif, dapat dilihat pada persamaan (16).

$$Presisi = \frac{TP}{(TP + FP)} \quad (16)$$

Recall atau sensitivitas adalah ukuran kinerja yang menunjukkan seberapa baik suatu pengklasifikasi dalam mengklasifikasikan kelas positif. *Recall* adalah rasio antara jumlah observasi positif yang diklasifikasikan secara benar dengan jumlah observasi positif asli, dapat dilihat pada persamaan (17).

$$Recall = \frac{TP}{(TP + FN)} \quad (17)$$

F1-Score adalah *harmonic mean* antara *precision* dan *recall*, dapat dilihat pada persamaan (18).

$$F1 = 2 * \frac{Precision * Recall}{(Precision + Recall)} \quad (18)$$

2.2 State of the Art

Menurut (Devi, 2017), Penyaluran minyak dan gas bumi dengan pipa bawah laut sangat efisien karena memerlukan biaya yang minimal. Namun seiring padatnya aktivitas maritim oleh lalu lintas kapal akan berdampak kerusakan pada jaringan pipa bawah laut. Penyebab kerusakan tersebut dapat disebabkan oleh beberapa faktor, antara lain: jatuhnya jangkar kapal (*anchor drop*), kapal tenggelam akibat gagal mesin, jaring atau pukat (*trawl*), dan faktor lainnya. Oleh karena itu, penilaian risiko untuk kapal

yang melewati jaringan pipa bawah laut sangat diperlukan. Pada penelitian ini, metode *Bayesian Network* (BN) digunakan untuk memodelkan penyebab kecelakaan pipa bawah laut oleh faktor kapal dan menghitung probabilitas kerusakan yang ditimbulkannya. Adapun standar DNV RP F107 digunakan untuk mengklasifikasi tingkat risiko berdasarkan nilai probabilitas yang didapat. Hasil pengujian program ini menunjukkan bahwa 58.4% kemungkinan kapal yang lewat tidak menyebabkan kerusakan pada pipa, 13.83% kerusakan yang ditimbulkan kecil, 15.14% kerusakan yang ditimbulkan menengah, dan 12.59% kerusakan yang ditimbulkan besar.

Menurut (Chow & Liu, 1968), sebuah metode yang ditampilkan untuk memperkirakan secara optimal distribusi probabilitas diskrit dimensi- n oleh hasil distribusi orde 2 atau distribusi pohon ketergantungan orde pertama. Masalahnya adalah untuk menemukan himpunan optimal $n-1$ hubungan ketergantungan orde pertama di antara n variabel, ditunjukkan bahwa prosedur yang diperoleh dalam makalah ini menghasilkan perkiraan perbedaan informasi yang kecil. Lebih lanjut ditunjukkan bahwa Ketika prosedur ini diterapkan pada pengamatan empiris dari distribusi *dependence tree* yang tidak diketahui prosedur tersebut adalah perkiraan kemungkinan maksimum dari distribusi tersebut.

Menurut (Cano, Sordo, & Gutierrez, 2004), Dalam tulisan ini ditampilkan beberapa aplikasi *Bayesian Network* dalam meteorologi dari sudut pandang data mining. Mereka bekerja dengan database observasi (curah hujan harian dan kecepatan angin maksimum) di jaringan 100 stasiun di semenanjung Iberia dan dengan pola atmosfer grid yang sesuai yang dihasilkan oleh model sirkulasi numerik. Sebagai langkah pertama, mereka menganalisis efisiensi algoritme pembelajaran standar untuk mendapatkan grafik asiklik terarah yang mewakili dependensi spasial di antara variabel yang termasuk dalam database, mereka juga menyajikan algoritma pembelajaran lokal baru yang memanfaatkan karakter spasial masalah. Model grafis yang dihasilkan diterapkan pada masalah meteorologi yang berbeda termasuk ramalan cuaca dan pembuatan cuaca stokastik.

Menurut (Verma & Dutta, 2018), makalah ini menampilkan teknik klasifikasi untuk menganalisis dataset jamur. Jamur dataset buatan terdiri dari catatan berbagai jenis jamur tidak beracun atau tidak bisa dimakan. *Artificial Neural Network* dan *Adaptive Nuero Fuzzy Inference System* digunakan untuk Teknik implementasi klasifikasi. Teknik yang berbeda digunakan untuk klasifikasi seperti ANN, ANFIS, dan *Naïve Bayes* digunakan mengkategorikan jamur yang berbeda sebagai kategori yang dapat dimakan atau tidak dapat dimakan. Performa dari teknik yang berbeda dievaluasi menggunakan akurasi, MAE, *Kappa Statistic*. Setelah menganalisa hasil, ditemukan bahwa *Adaptive Nuero Fuzzy inference System* mengungguli teknik lain dengan akurasi tertinggi, kesalahan absolut rata-rata terendah dan ANN adalah yang berkinerja terbaik kedua. Jika ukuran set pelatihan ditingkatkan, akurasi juga meningkat dengan menghormati set pelatihan.

Menurut (Hruschka Jr, Hruschka, & Ebecken, 2007), Nilai yang hilang adalah masalah penting dalam penambahan data untuk mengatasi masalah klasifikasi, mereka mengusulkan dua metode imputasi berdasarkan *Bayesian Network*. Metode ini dievaluasi baik dalam konteks prediksi maupun klasifikasi. Mereka membandingkan hasil yang diperoleh dengan yang dicapai dengan metode imputasi klasik (*Expectation–Maximization, Data Augmentation, Decision Trees, and Mean/Mode*). Simulasi dilakukan melalui empat kumpulan data (*Congressional Voting Records, Mushroom, Wisconsin Breast Cancer and Adult*) melakukan tolak ukur untuk metode data mining. Nilai yang hilang disimulasikan dalam kumpulan data dengan cara eliminasi beberapa nilai yang diketahui. Dengan demikian, dimungkinkan untuk menilai kemampuan prediksi dari sebuah metode imputasi, membandingkan nilai asli dengan yang diperhitungkan. Selain itu, kami mengusulkan metodologi untuk memperkirakan bias yang disisipkan oleh metode imputasi dalam klasifikasi.

2.3 Kerangka Konseptual

Pendahuluan: Diperkirakan terdapat 1,5 juta spesies jamur di dunia, jenis yang teridentifikasi sebanyak 28.700 jenis jamur makroskopis (memiliki tubuh buah), jamur mikroskopik (tidak memiliki tubuh buah) sebanyak 24.000 dan 13.500 jenis lumut kerak (asosiasi simbiotik antara fungi dan alga), sedangkan jenis jamur yang belum teridentifikasi sejumlah 1.433.800 jenis, baik makro maupun mikro.



Masalah: Untuk membedakan jamur *family Agaricus* dan *Lepiota* tidak beracun dan yang beracun didasarkan pada bentuk, sifat, dan keadaannya, sangat sukar dilakukan. Hal ini dikarenakan adanya bentuk hampir sama dari spesies jamur *family Agaricus* dan *Lepiota* yang dapat di konsumsi dengan spesies jamur lain yang beracun.



Solusi: Beberapa penelitian telah dilakukan dengan berbagai metode berbeda, *naïve Bayes* memiliki akurasi sebesar 99,85%, *voting feature interval* memiliki akurasi sebesar 84,53%, *artificial neural network* memiliki akurasi sebesar 70%, *adaptive neural fuzzy inference* memiliki akurasi 80%.



Metode: *Bayesian Network* ini dapat mengeksploitasi hubungan bebas bersyarat (*conditional independence*) dalam membangun struktur jaringan, sehingga dapat membangun sebuah model yang lebih tersusun dan mengurangi kompleksitas perhitungan dalam melakukan inferensi.



Hasil: Dengan adanya penelitian ini, diharapkan dapat mengklasifikasi jamur beracun dan tidak beracun.