



**IMPLEMENTASI METODE NAÏVE BAYES CLASSIFIER
UNTUK IDENTIFIKASI JENIS JAMUR**

Septian AriePrayoga¹, Ismasari Nawangsih², Tri NgudiWiyatno³

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
Korespondensi email: ismasari.n@pelitabangsa.ac.id

Abstract

Being in a country that has abundant natural resources and biodiversity, of course, something that should be grateful for, such as its diversity of flora. The climate is very suitable for growing various types of plants. The climate is very suitable for growing various types of plants. One of them is the food sector, especially horticulture, a horticultural commodity that is often found and widely used consumption materials by the community are mushrooms. The types of fungi of the genus Agaricus and Lepiota are often found around the community. There are types of mushrooms that are edible and some are poisonous. The type of fungi genus agaricus and lepiota which have poisons can cause pain in people who consume them, can even have an impact on death. To distinguish between mushrooms that are safe for consumption and those that are poisonous is very difficult. Therefore the need for problem solving to contribute to identifying the fungi is safe to eat or toxic. In this study the researcher tried to contribute in solving the problem in order to identify fungi using the naive bayes classifier method. Naive bayes classifier will classify the fungi dataset of genus agaricus and lepiota using probability calculations. In the scheme the testing data will be inputted with certain attributes, with subsequent training data with the NBC method testing data will be classified with training data so that the results of the testing data are included in the edible or poisonous category. The data used in this study is a collection of mushroom data (genus agaricus and genus lepiota), this dataset is taken from the UCI Machine learning repository. The mushroom data has 23 attributes based on morphological characteristics, with a total of 8124 mushroom records. The mushroom data has 23 attributes based on morphological characteristics, with a total of 8124 mushroom records. After using the 10-fold cross validation scheme, the average method of naive bayes method in the history of agaricus and lepiota mushrooms was found to be an accuracy value of 86.64 using the method of good classification.

Informasi Artikel

Diterima: 2 Sept 2019
Direvisi: 5 Sept 2019
Dipublikasikan:
9 Sept 2019

Keywords

Fungi, Agaricus,
Lepiota, Naive bayes
classifier

I. Pendahuluan

Berada di dalam sebuah negara yang memiliki kekayaan alam serta keanekaragaman hayati yang berlimpah tentunya sesuatu hal yang patut disyukuri, seperti keanekaragaman floranya. Iklim yang sangat cocok untuk tumbuhnya berbagai macam jenis tumbuhan. Salah satu yang dinilai mempunyai prospek cerah adalah komoditas pertanian. Perkembangan pertanian di bidang pangan khususnya *hortikultura* pada sekarang ini ditujukan untuk lebih memantapkan swasembada pangan, meningkatkan pendapatan masyarakat dan diharapkan dapat memperbaiki keadaan gizi masyarakat. Salah satu komoditas *hortikultura* yang sering dijumpai dan banyak dijadikan bahan konsumsi oleh masyarakat adalah jamur. Jamur bisa dijadikan sumber pangan bagi masyarakat dan juga dapat digunakan sebagai bahan obat (Ian R.Hall et al., 2003).

Seiring dengan perkembangan jaman, telah diketahui terdapat lebih dari ribuan jamur dengan berbagai jenis, mulai dari wilayah subtropis dengan suhu dingin hingga wilayah tropis dengan suhu hangat. Namun dari ribuan jenis jamur yang tersebar tersebut terdapat jamur yang beracun dan jamur yang aman dikonsumsi. Jamur dapat tumbuh dengan mudah di alam bebas (Sysouphanthong et al., 2011). Salah satu jenis jamur yang banyak dijumpai di alam, dibudidayakan atau dikonsumsi adalah jenis jamur yang masuk kedalam genus *agaricus* dan *lepiota*. Genus *agaricus* dan *lepiota* dapat hidup di alam yang terbuka, datang dalam berbagai warna, bentuk, tekstur dan ukuran yang masih banyak belum diketahui, ada yang beracun ada juga yang aman (Ian R.Hall et al., 2003). Dan hampir setiap jenis jamur memiliki ciri-ciri yang hampir serupa (Ian R.Hall et al., 2003). Dengan itu tentunya sangat sulit membedakan mana jamur yang beracun dan aman jika tidak cukup pengetahuan tentang jamur. Begitu banyak kejadian keracunan makanan akibat jamur, dikarenakan tidak mengetahui bahwa jamur tersebut beracun. Bahkan sejumlah negara melaporkan kasus keracunan akibat jamur beracun.

Untuk membedakan Jamur konsumsi dan beracun penulis mencoba menggunakan suatu teknik klasifikasi untuk membedakannya yaitu, Metode *Naïve Bayes Classifier* dengan menggunakan data penelitian yang telah ada. Metode *Naive Bayes Classifier* akan menghasilkan jamur masuk kedalam kelas aman dikonsumsi atau beracun. Algoritma ini mengasumsikan bahwa atribut bersifat independen (Markus et al., 2014).

Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai teorema bayes.

Data yang digunakan pada penelitian ini adalah dataset jamur dalam genus *agaricus* dan *lepiota* yang diambil dari *UCI Machine Learning Repository*, dataset tersebut digambarkan dengan 22 atribut yang berbeda (ciri morfologi) dengan kelas konsumsi dan beracun, serta terdapat 8000 lebih jumlah data (UCI Repository, 1987).

1.1 Rumusan Masalah

- 1 Bagaimana hasil tingkat akurasi yang didapatkan oleh perhitungan metode *Naive Bayes Classifier* dalam menentukan jenis jamur konsumsi dan jamur beracun ?
- 2 Bagaimanana rancang aplikasi untuk mengklasifikasikan jamur konsumsi dan jamur beracun menggunakan metode *Naive Bayes Classifier* ke dalam kelas, konsumsi (*edible*) dan beracun (*poisonous*) ?

1.2 Tujuan Penelitian

- 1 Mengetahui tingkat akurasi hasil keluaran yang diberikan perhitungan metode *Naive Bayes Classifier* dalam mengklasifikasikan jamur konsumsi dan beracun, sebagai acuan atau tolak ukur seberapa baik metode tersebut berhasil diterapkan dalam mengklasifikasikan jamur.
- 2 Merancang aplikasi untuk mengklasifikasikan jamur konsumsi dan jamur beracun menggunakan metode *Naive Bayes Classifier*.

2. METODOLOGI PENELITIAN

A. DATASET

Data didapatkan dari UCI *Machine Learning Repository*, UCI merupakan kumpulan basis data, yang digunakan oleh komunitas *machine learning* untuk analisis empiris algoritma *machine learning*. Data tersebut merupakan

data penelitian yang sudah dilakukan oleh peneliti sebelumnya, dengan jumlah data ciri-ciri jamur (Morfologi) tergolong konsumsi dan beracun sebanyak 8124 *record*. Berikut adalah bentuk atribut ciri-ciri jamur konsumsi dan beracun beserta dataset yang digunakan yang didapat dari peneliti sebelumnya :

TABEL 1. ATRIBUT JAMUR

No	Atribut	Karakteristik Atribut
1	cap-shape	Kategorikal
2	cap-surface	Kategorikal
3	cap-color	Kategorikal
4	Bruises	Kategorikal
5	Odor	Kategorikal
6	gill-attachment	Kategorikal
7	gill-spacing	Kategorikal
8	gill-size	Kategorikal
9	gill-color	Kategorikal
10	stalk-shape	Kategorikal
11	stalk-root	Kategorikal
12	stalk-surface-above-ring	Kategorikal
13	stalk-surface-below-ring	Kategorikal
14	stalk-color-above-ring	Kategorikal
15	stalk-color-below-ring	Kategorikal
16	veil-type	Kategorikal
17	veil-color	Kategorikal
18	ring-number	Kategorikal
19	ring-type	Kategorikal
20	spore-print-color	Kategorikal
21	Population	Kategorikal
22	Habitat	Kategorikal
23	edible / poisonous	Kategorikal

TABEL 2. DATASET JAMUR

X	s	N	t	p	f	C	n	k	e	e	s	s	w	W	p	W	o	P	k	s	u	p
X	s	Y	t	a	f	C	b	k	e	c	s	s	w	W	p	W	o	P	n	n	g	e
B	s	W	t	i	f	C	b	n	e	c	s	s	w	W	p	W	o	P	n	n	m	e
X	y	W	t	p	f	C	n	n	e	e	s	s	w	W	p	W	o	P	k	s	u	p
X	s	G	f	n	f	W	b	k	t	e	s	s	w	W	p	W	o	E	n	a	g	e
X	y	Y	t	a	f	C	b	n	e	c	s	s	w	W	p	W	o	P	k	n	g	e
B	s	W	t	a	f	C	b	g	e	c	s	s	w	W	p	W	o	P	k	n	m	e
B	y	W	t	l	f	C	b	n	e	c	s	s	w	W	p	W	o	P	n	s	m	e
X	y	W	t	p	f	C	n	p	e	e	s	s	w	W	p	W	o	P	k	v	g	p

Dataset pada Tabel 2 adalah gambaran dataset dari keseluruhan dataset jamur sebanyak 8124 *record*. Dengan gambaran *horizontal* adalah atribut, *vertikal* adalah banyaknya dataset yang tergolong dapat diamankan dan beracun yaitu 8124 *record*.

B. DATA PRE-PROCESSING

Pada penelitian ini akan dilakukan penggantian dataset sesuai informasi yang dipublish oleh peneliti sebelumnya tentang arti dari nilai yang terdapat pada setiap atribut agar mudah dimengerti dan juga akan dilakukan proses penanganan *missing values* pada dataset jamur, ini dilakukan agar menghasilkan kualitas data

masukannya yang tidak menimbulkan kesalahan pada saat *mining*.

i) *Missing Values*

Missing values merupakan informasi yang tidak tersedia untuk sebuah objek (kasus). *Missing values* terjadi karena informasi untuk sesuatu tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak

ada. Penanganan *missing values* pada penelitian ini karena dataset yang akan digunakan mengandung *missing values*. *Missing values* pada dasarnya tidak bermasalah bagi keseluruhan data, toleransi *missing values* dari keseluruhan data sebesar 5 % namun jika presentase data yang hilang tersebut cukup besar, maka diperlukan penanganan terhadap data yang *missing* (Daniel et al., 2014). Beberapa teknik untuk mengganti nilai yang hilang adalah sebagai berikut (Daniel et al, 2014) :

- Ganti nilai yang hilang dengan beberapa konstanta, yang ditentukan oleh analisis. Atau membuang nilai yang hilang.
- Ganti nilai yang hilang dengan *mean* (untuk variabel numerik) atau (untuk variabel kategori).
- Ganti nilai yang hilang dengan nilai yang dihasilkan secara acak dari pengamatan distribusi variabel.

- Ganti nilai-nilai yang hilang dengan nilai yang dihitung berdasarkan karakteristik lain dari catatan.

ii) Transformasi Data

Terdapat 2 kategori *Missing value* yaitu, MV Nilai Atribut Terhadap Kelas dan MV seluruh *record* atribut. Pada MV nilai atribut data akan disisihkan jika tidak sama dengan dataset publisher dan pada MV seluruh *record* atribut terlebih dahulu ditetapkan ketentuan transformasi, dalam penelitian ini adalah mengganti nilai atribut yang berbentuk kategorikal ke dalam bentuk numerik pada semua atribut sebanyak 23 atribut. Nilai numerik di dapat berdasarkan banyaknya nilai atribut. Penggunaan *mean* menganjurkan nilai dalam bentuk numerik. Berikut sebagian contoh ketentuan transformasi kategorikal ke numerik, jika salah satu dari atribut mengandung *missing*:

TABEL 3. TRANSFORMASI DATASET KATEGORIKAL KE NUMERIK

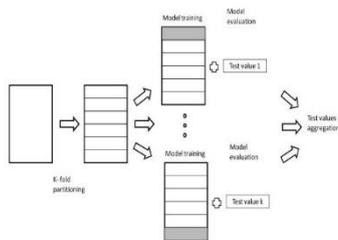
No	Atribut	Kategorikal	Numerik
1	Bentuk Tudung	<ul style="list-style-type: none"> ▪ Bell ▪ Conical ▪ Convex ▪ Flat ▪ Knobbed ▪ Sunken 	<ul style="list-style-type: none"> ▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 5 ▪ 6
2	Permukaan Tudung	<ul style="list-style-type: none"> ▪ Fibrous ▪ Grooves ▪ Scaly ▪ Smooth 	<ul style="list-style-type: none"> ▪ 1 ▪ 2 ▪ 3 ▪ 4
...	dst.....	▪ dst.....	▪ dst.....
4	Terdapat Memar	<ul style="list-style-type: none"> ▪ Bruises ▪ No 	<ul style="list-style-type: none"> ▪ 1 ▪ 2
...	dst.....	▪ dst.....	▪ dst.....
11	Bentuk Akar Tangkai	<ul style="list-style-type: none"> ▪ Bulbous ▪ Club ▪ Cup ▪ Equal ▪ Rhizomorphs ▪ Rooted 	<ul style="list-style-type: none"> ▪ 1 ▪ 2 ▪ 3 ▪ 4 ▪ 5 ▪ 6
...	dst.....	▪ dst.....	▪ dst.....
...	dst.....	▪ dst.....	▪ dst.....
23	Kategori	<ul style="list-style-type: none"> ▪ Edible ▪ Poisonous 	<ul style="list-style-type: none"> ▪ 1 ▪ 2

C. K-FOLD CROSS VALIDATION

Dalam *k-fold cross validation*, data asli dipartisi kedalam subset/bagian. Model ini kemudian dibangun menggunakan data dari subset K-1 (2,3,4,5, dst), dan terdapat bagian

lain didalam subset yang digunakan untuk set tes. Bagian subset (dataset) harus lebih banyak dari pada set tes, dilakukan secara iteratif sampai memiliki model yang berbeda. Hasil dari masing-masing model K (akurasi) kemudian digabungkan menggunakan rata-rata untuk mendapatkan hasil akurasi dari

keseluruhan data. *K-fold cross-validation* yang sering digunakan adalah *10-fold cross validation*. Manfaat menggunakan *k-fold cross-validation* adalah setiap record muncul dalam set tes tepat satu kali, kekurangannya adalah bahwa tugas validasi yang diperlukan dibuat lebih sulit (Daniel et al, 2014).



GAMBAR 1. Proses K-Fold Cross Validation

D. NAÏVE BAYES CLASSIFIER

Salah satu algoritma klasifikasi probabilistik, algoritma *naive bayes* merupakan penggolongan probabilistik sederhana berdasarkan penerapan teorema bayes dengan asumsi independensi yang kuat, dengan kata lain algoritma *naive bayes* mengasumsikan bahwa keberadaan nilai tertentu dari suatu atribut tidak terkait dengan kehadiran nilai atribut lainnya (Markus et al., 2014). Klasifikasi *naive bayes* sangat cocok ketika dimensi input tinggi, dan memiliki kinerja yang sebanding dengan beberapa metode klasifikasi lainnya seperti *decision tree* dan *neural network classifier* (Charu, 2015). Berikut adalah persamaan dari teori naive bayes :

$$p(H|D) = \frac{p(D)p(D|H)}{p(D)}$$

Keterangan :

- P : Probabilitas
- D : Data dengan class yang belum diketahui.
- H : Hipotesis data merupakan class spesifik.
- $p(H|D)$: Probabilitashipotesis H berdasarkan kondisi D (posterior probability).
- $p(H)$: Probabilitas hipotesis H (prior probability).

$p(D|H)$: Probabilitas D berdasarkan kondisi pada hipotesis H.
 $p(D)$: Probabilitas D.

E. CONFUSION MATRIX

Evaluasi untuk mengukur kinerja model digunakan *confusion matrix*, karena *confusion matrix* merupakan alat yang berguna untuk menganalisa seberapa baik pengklasifikasi dapat mengenali tupel/fitur dari kelas yang berbeda (Han, Kamber, & Pei, 2011). *Confusion matrix* dapat membantu menunjukkan rincian kinerja pengklasifikasi dengan memberikan informasi jumlah fitur suatu kelas yang diklasifikasikan dengan tepat dan tidak tepat. *Confusion matrix* memberikan penilaian kinerja model klasifikasi berdasarkan jumlah objek yang diprediksi dengan benar dan salah. Pengukuran akurasi dengan *confusion matrix* merupakan matrik 2 dimensi yang meng- gambarkan perbandingan antara hasil prediksi dengan kenyataan.

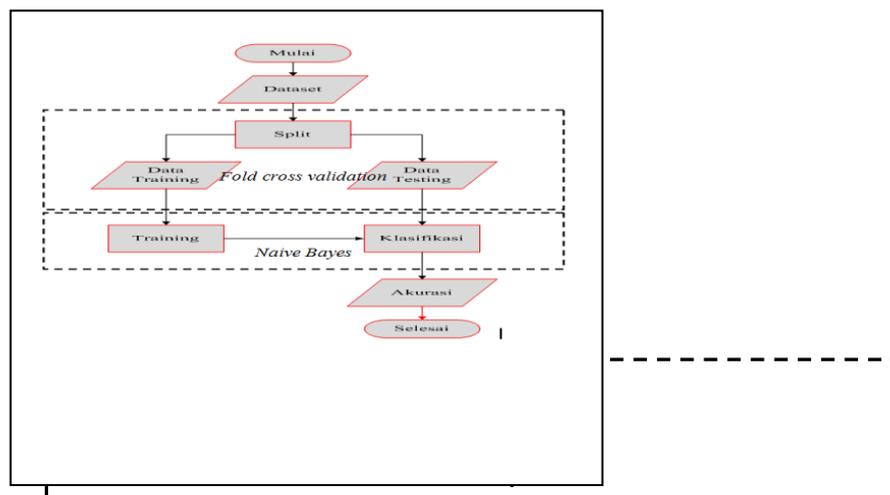
TABEL 4. CONFUSION MATRIX

Confusion Matrix		Class Actual	
		Positive	Negative
Class Predicted	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

F. PROSES PENGUJIAN

Mengevaluasi kinerja metode klasifikasi umumnya menggunakan hasil keseluruhan pada pengujian dataset. Pengujian kinerja pada

metode *naive bayes* akan dilakukan dengan menggunakan *k-fold cross validation*. Metode evaluasi yang umum digunakan yaitu *10-fold cross validation*, *10-fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian.



GAMBAR 2. Flow chart Pengujian

i) *Fold Cross Validation*

- Dataset, yang digunakan pada tahap pengujian ini adalah dataset jamur sebanyak 8124 *record* yang sudah diolah pada tahap *preprocessing*, sehingga data sudah layak untuk diolah lebih lanjut.
- Split, Pembagian data *training* dan data *testing* pada dataset jamur menggunakan metode *10-fold cross validation*, dataset akan dibagi menjadi 10 bagian dari 8124 data jamur kategori dapat dimakan dan beracun dengan proporsi yang hampir sama. Satu bagian (data *testing*) akan digunakan untuk pengujian, sedangkan

bagian-bagian lainnya (data *training*) akan digunakan untuk pelatihan. Dengan ketentuan data *training* lebih banyak proporsinya.

ii) *Naive Bayes*

- *Training* merupakan penentuan data yang akan dijadikan input untuk untuk pengujian menggunakan metode *naive bayes*. Data yang akan dijadikan input adalah data *testing* yang sudah dibagi sama rata. Data tersebut nantinya akan diuji berdasarkan data *training* yang juga sudah dibagi sama rata.

TABEL 5.SKEMA TRAINING

ID Data Training	Jumlah Data Training	ID Data Testing	Jumlah Data Testing
DTR-1	7311 Data	K-2, K-3, K-4, K-5, K-6, K-7, K-8, K-9, k-10	812 Data
DTR-2	7311 Data	K-1, K-3, K-4, K-5, K-6, K-7, K-8, K-9, K-10	812 Data
DTR-3	7311 Data	K-1, K-2, K-4, K-5, K-6, K-7, K-8, K-9, K-10	812 Data
DTR-4	7311 Data	K-1, K-2, K-3, K-5, K-6, K-7, K-8, K-9, K-10	812 Data
DTR-5	7312 Data	K-1, K-2, K-3, K-4, K-6 K-7, K-8, K-9, K-10	812 Data
...	dst.....	dst.....	dst.....
...	dst.....	dst.....	dst.....
...	dst.....	dst.....	dst.....
...	dst.....	dst.....	dst.....
DTR-10	7312 Data	K-1, K-2, K-3, K-4, K-5 K-6, K-7, K-8, K-9	813 Data

- **Klasifikasi**

Dalam tahap ini akan dilakukan klasifikasi/pengujian data *testing* terhadap data *training* yang sudah dibagi berdasarkan metode *fold cross validation* dengan perhitungan *naive bayes*, pada tahap ini juga akan diketahui akurasi dari masing-masing pengujian.

TABEL 6. SKEMA TRAINING

ID Pengujian	ID Data Training	ID Data Testing
PJN-1	DTR-1	K-1
PJN-2	DTR-2	K-2
PJN-3	DTR-3	K-3
PJN-4	DTR-4	K-4
PJN-5	DTR-5	K-5
PJN-6	DTR-6	K-6
PJN-7	DTR-7	K-7
PJN-8	DTR-8	K-8
PJN-9	DTR-9	K-9
PJN-10	DTR-10	K-10

- **Akurasi**

Pada tahap ini akan dilakukan perhitungan akurasi menggunakan *tools* rapid miner pada setiap pengujian berdasarkan *confusion matrix* masing-masing pengujian, terdapat 10 pengujian. Adapun rumus dari akurasi berdasarkan *confusion matrix* adalah sebagai berikut :

$$\text{Akurasi Setiap Pengujian} = \frac{TP+TN}{(TP+FN)+(FP+TN)} \times 100$$

Selanjutnya akan dilakukan perhitungan rata-rata akurasi berdasarkan keseluruhan pengujian. Ini dilakukan untuk mengetahui

seberapa besar akurasi dari keseluruhan data, sebagai acuan seberapa baik metode *naive bayes* dalam mengidentifikasi jamur. Tingkat akurasi menunjukkan tingkat kebenaran klasifikasi terhadap kelas. Semakin rendah akurasi, semakin tinggi kesalahan dan sebaliknya semakin tinggi akurasi maka semakin rendah kesalahan. Akurasi yang baik adalah akurasi yang mendekati 100%. Perhitungan akurasi dapat dilakukan sebagai berikut :

$$\text{Rata-Rata Akurasi} = \frac{\sum \text{Hasil Akurasi setiap pengujian}}{\sum \text{Jumlah Total Pengujian}}$$

3. HASIL DAN PEMBAHASAN

G. HASIL DATA PRE-PROCESSING

Pada bagian ini akan dijelaskan berbagai hal yang berkaitan dengan hasil *pre-processing* data dengan tool SPSS, hasil dari perhitungan *naive bayes* dengan menggunakan pengujian metode *10-fold cross validation* pada tool rapid miner.

i) Hasil Penanganan *Missing Values* Nilai Atribut Terhadap Kelas Setelah diamati terdapat beberapa atribut yang memiliki nilai atribut kosong terhadap kelas *edible* dan *poisonous*. Meliputi atribut Bentuk Bilah, Jarak Bilah, Tipe Tudung, Tipe Cincin, Bentuk Akar Tangkai. Berikut gambaran kecilnya :

TABEL 7. PERBANDINGAN NILAI ATRIBUT BENTUK BILAH DENGAN HASIL SPSS

Bentuk Akar Tangkai (Peneliti Sebelumnya)	Hasil SPSS
BULBOUS	BULBOUS
CLUB	CLUB
CUP	-
EQUAL	EQUAL
RHIZOMORPHS	-
ROOTED	ROOTED

Dengan demikian nilai atribut yang tidak terdapat pada hasil output SPSS akan disisihkan atau tidak digunakan. Dikarenakan nilai atribut tersebut tidak memiliki kontribusi terhadap kelas *edible* maupun *poisonous*.

ii) Hasil Penanganan *Missing Values* Atribut Jamur.

Setelah dilakukan pengolahan menggunakan SPSS, Terlihat bahwa atribut yang memiliki *missing values* hanya atribut Bentuk Akar Tangkai dengan total data pada atribut tersebut sebanyak 8124, dengan jumlah data *valid* sebanyak 5644 data, dan data yang *missing* sebanyak 2480 dengan besaran persentase *missing* adalah 30,5 %. *Missing* yang terdapat pada pada atribut tersebut persentasenya cukup besar.

TABEL 8.HASIL PENCARIAN MISSING VALUES ATRIBUT JAMUR

Univariate Statistics			
	N	Missing	
		Count	Percent
Bentuk_Tudung	8124	0	,0
.....	dst...	dst...	dst...
.....	dst...	dst...	dst...
Bentuk_Tangkai	8124	0	,0
Bentuk_Akar_Tangkai	5644	2480	30,5
Permukaan_Tangkai_Diatas_Cincin	8124	0	,0
.....	dst...	dst...	dst...
.....	dst...	dst...	dst...
Habitat	8124	0	,0
Kategori	8124	0	,0

Atribut yang mengandung *missing* dalam hal ini adalah atribut Bentuk Akar Tangkai akan diolah. Perlu diketahui terlebih dahulu nilai atribut CUP dan RHIZOMORPHS akan disisihkanserta untuk menangani *missing values* terlebih dahulu merubah data ke bentuk numerik seperti ketentuan yang sudah dijelaskan pada Tabel 3.

TABEL 9.HASIL ATRIBUT BENTUK AKAR TANGKAI TANPA MISSING VALUES

SMEAN(Bentuk_Akar_Tangkai_1)					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	3776	46,5	46,5	46,5
	1,86	2480	30,5	30,5	77,0
	2,00	556	6,8	6,8	83,9
	4,00	1120	13,8	13,8	97,6
	6,00	192	2,4	2,4	100,0
Total		8124	100,0	100,0	

Setelah dilakukan penanganan *missing value* soleh SPSS dengan *mean*, keterangan *missing values* sudah digantikan dengan nilai 1,86. Nilai 1,86 dengan demikian mendefinisikan nilai yang dibulatkan menjadi 2 dengan artian *missing values* akan diisi dengan nilai atribut kategorikal CLUB.

H. PENGUJIAN

Telah dijelaskan bahwa proses pengujian, menggunakan perhitungan *naive bayes* dengan skema *10-fold cross validation*, dengan artian dataset akan dibagi menjadi N bagian secara acak. Fold K-1 adalah ketika bagian ke-1 menjadi data *testing* dan sisanya menjadi data *training*, demikian seterusnya hingga sampai Fold K-10 bagian ke-10. Pengujian pada penelitian ini menggunakan *tool* rapid miner.

i) Hasil Pengujian

Pengujian dilakukan

sesuai dengan yang telah tertuang pada Tabel 6. Berikut adalah hasil pengujian berupa *confusion matrix* berdasarkan perhitungan metode *naive bayes classifier* berdasarkan skema *10-fold cross validation*

accuracy: 89.67%			
	true POISONOUS	true EDIBLE	class precision
pred. POISONOUS	0	0	0.00%
pred. EDIBLE	84	729	89.67%
class recall	0.00%	100.00%	

Gambar 3. Hasil Pengujian 1 (DTR 1 – K 1)

TABEL 10. KLASIFIKASI BENAR PENGUJIAN 1

Jamur <i>Agaricus</i> dan <i>Lepiota</i>		Hasil		<i>Confusion Matrix</i>
Data Testing	Klasifikasi Benar	Actual	Prediksi	<i>True Accuracy</i>
813 Data	729 Data	EDIBLE	EDIBLE	$\text{Akurasi} = \frac{0+729}{(0+0)+(84+729)} \times 100$ $= \frac{729}{813}$ $= 89,67 \%$

Pada Gambar 3. Hasil Pengujian 1 dengan data testing sebanyak 813 data dan data training sebanyak 7311 data, terlihat dari *confusion matrix* bahwa terdapat 84

data testing dengan kategori poisonous, diklasifikasikan sebagai edible oleh model tetapi aktualnya adalah poisonous dengan demikian model salah dalam mengklasifikasikan, dan sebanyak 729 data testing dengan kategori edible diklasifikasikan sebagai edible, dengan demikian model benar dalam mengklasifikasikan dan didapati akurasi sebesar 89,67%. Pengujian itu dilakukan sampai pengujian ke 10. Rincian Hasil Per Rincian Hasil Perhitungan *Naïve Bayes Classifier*

TABEL 11. RINCIAN HASIL KLASIFIKASI JAMUR AGARICUS DAN LEPIOTA

Pengujian	Data Training	Data Testing	Klasifikasi Benar (Tes)	Klasifikasi Salah (Tes)	Acc True	Acc False
Pengujian 1	7311	813	729 Data	84 Data	89,67%	10,33%
Pengujian 2	7311	813	732 Data	81 Data	90,04%	9,96%
Pengujian 3	7311	813	724 Data	89 Data	89,05%	10,95%
Pengujian 4	7311	813	656 Data	157 Data	80,69%	19,31%
Pengujian 5	7312	812	589 Data	223 Data	72,54%	27,46%
Pengujian 6	7312	812	736 Data	76 Data	90,64%	9,36%
Pengujian 7	7312	812	556 Data	256 Data	68,47%	31,53%
Pengujian 8	7312	812	704 Data	108 Data	86,70%	13,30%
Pengujian 9	7312	812	807 Data	5 Data	99,38%	0,62%
Pengujian 10	7312	812	756 Data	56 Data	93,10%	6,90%

Pada Tabel 11 dapat dilihat pada pengujian 1 sampai 3 memiliki *trend* akurasi yang relatif stabil sedangkan dari pengujian 4 *accuracy* yang fluktuasi. Dari hasil pengujian bahwa data *training* dan data *testing* yang berbeda menghasilkan akurasi yang berbeda pula disebabkan karena variasi dan jumlah nilai atribut pada dataset yang berbeda-beda.

sampai 10 memiliki *trend* akurasi turun naik. Ini menandakan dari setiap pengujian bahwasanya memiliki

ii) Hasil Akhir Seluruh Pengujian Hasil rata-rata akurasi ini akan dijadikan acuan seberapa baik metode *naive bayes* diterapkan dalam mengidentifikasi jamur. Maka dari masing-masing pengujian didapatkan hasil rata-rata akurasi perhitungan *naive bayes classifier*, sebagai berikut:

$$\text{Rata-rata Akurasi} = \frac{89,67 + 90,04 + 89,05 + 80,69 + 72,54 + 90,64 + 68,47 + 86,70 + 99,38 + 93,10}{10}$$

$$= \frac{860,64}{10} = 86,06 \%$$

Berdasarkan dari pengujian dan setelah dihitung rata-rata akurasi dari setiap pengujian didapatkan hasil akhir akurasi sebesar 86,06%. Ini mendefinisikan dalam penelitian ini, bahwa metode *naive bayes* termasuk *good classification*.

4. KESIMPULAN

Akurasi pada setiap pengujian (Pengujian 1 – Pengujian 10) mengalami fluktuasi, dari hasil pengujian bahwa data *training* dan data *testing* yang berbeda menghasilkan akurasi yang berbeda pula disebabkan karena variasi dan jumlah nilai atribut pada dataset yang berbeda-beda. Sehingga akurasi yang didapat dari keseluruhan pengujian untuk menentukan jenis jamur konsumsi dan jamur beracun didapati hasil 86,06%, tergolong kedalam *good classification*. Menunjukkan bahwa metode *naive bayes classifier* baik diterapkan dalam identifikasi jamur *agaricus* dan *lepiota* kategori *edible* atau *poisonous*.

5. DAFTAR PUSTAKA

- Aggarwal, Charu C. 2014. *Data Classification Algorithms And Applications*. Chapman And Hall/CRC Data Mining and Knowledge Discovery Series. United States : CRC Press.
- Hall, Ian R., Stephenson, Steven L., Buchanan, Peter K., Yun, Wang., Cole, Anthony L.J. 2003. *Edible And Poisonous Mushrooms Of The World*. Portland Cambridge : Timber Press.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. San Francisco: Morgan Kaufmann Publishers Inc.
- Hofmann, Markus., Klinkenberg, Ralf. 2014. *Rapid Miner Data Mining Use Cases And Business Analytics Applications*. Chapman And Hall/CRC Data Mining and Knowledge Discovery Series. United States : CRC Press.
- Larose, Daniel.T., Larose, Chantal.D. 2014. *Discovering Knowledge In Data : An Introduction to Data Mining Second Edition*. Canada : John Wiley & Sons, Inc
- P, Sysouphanthong., Hyde K.D., E, Chukeatirote., Velinga. 2011. *A Review Of Genus Lepiota And its Distribution in East Asia*. Current Research in Environmental And Applied Mycology 1 (2) : 161-176.
- Schlimmer, Jeff. 1987. UCI Machine Learning Repository: *Mushroom Dataset*. Diakses 10 Juni 2018, dari <https://archive.ics.uci.edu/ml/datasets/mushroom>