

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DALAM MENENTUKAN KELAYAKAN KONSUMSI JAMUR

*Implementation Of K - Nearest Neighbor Algorithm In Determining
Feasibility of Fungus Consumption*

Johannes Christian, Muhamad Fadillah, Cindy Chairunnisa, Audrey Era Goldenia, Hilda Harisa
Program Studi Informatika
Universitas Pembangunan Nasional Veteran Jakarta

Abstrak

Jamur merupakan salah satu tumbuhan yang banyak dikonsumsi, namun dari berbagai jenis jamur yang ada, sulit untuk mengetahui apakah jamur tersebut layak dan dapat dikonsumsi atau tidak. Terdapat beberapa cara untuk mengetahui kelayakan jamur. Salah satu cara untuk mengatasi permasalahan tersebut adalah dengan melakukan klasifikasi menggunakan algoritma K-Nearest Neighbor (K-NN). Penerapan algoritma K-NN digunakan untuk menentukan kelayakan konsumsi dari suatu jamur. Hasil dari penelitian ini adalah klasifikasi jamur dapat dikonsumsi(edible) dan tidak dapat dikonsumsi(poisonous) dengan nilai akurasi yang diperoleh sebesar 99.01%.

Kata Kunci — Klasifikasi, Jamur, K-Nearest Neighbor (K-NN)

Abstract

Mushroom is one of the most consumed plants, but from the various types of mushrooms available, it is difficult to know whether the fungus is suitable and can be consumed or not. There are several ways to determine the feasibility of mushrooms. One way to overcome these problems is to do the classification using the K-Nearest Neighbor (K-NN) algorithm. The application of the K-NN algorithm is used to determine the feasibility of consumption of a mushroom. The results of this study are the classification of mushrooms can be consumed(edible) and cannot be consumed(poisonous) with an accuracy value obtained at 99.01%.

Keywords — Classification, Mushroom, K-Nearest Neighbor (K-NN)

1. PENDAHULUAN

1.1 Latar Belakang

Dalam kehidupan sehari-hari fungi atau yang biasa disebut jamur merupakan salah satu jenis makhluk hidup yang dapat dikonsumsi oleh manusia. Jamur yang tumbuh di alam seiring bertambahnya waktu semakin meningkat jumlahnya di mana jamur tersebut memiliki bentuk, warna serta jenis yang berbeda-beda. Namun, dari banyaknya jenis jamur yang tumbuh di alam ini tidak semuanya dapat dikonsumsi. Ada beberapa diantaranya merupakan jamur yang beracun dan tidak dapat dikonsumsi sehingga dapat memberikan efek buruk bagi kehidupan, seperti mual, diare, sakit perut, dan sebagainya.

Salah satu jenis jamur yang dapat dikonsumsi pada penelitian ini adalah jenis jamur *Agaricus Bisporus* dan *Lepiota Americana*. *Agaricus* merupakan salah satu genus dari jamur yang banyak tumbuh di alam terbuka yang di dalamnya terdiri dari jenis jamur yang dapat dikonsumsi dan tidak dapat dikonsumsi. Sehingga sulit untuk membedakan jenis jamur antara yang dapat dikonsumsi dan tidak hanya berdasarkan pada ciri dan bentuk dari jamur tersebut.

Lepiota merupakan genus jamur yang termasuk dalam famili *Agaricaceae*. Semua spesies jamur *Lepiota* adalah saprotrof yang hidup di darat dengan preferensi untuk tanah yang kaya dan berkapur. Jenis jamur ini terdapat sekitar 400 spesies yang saat ini diakui di seluruh dunia. Kebanyakan jenis jamur dalam spesies jamur ini merupakan jenis jamur beracun dan mematikan. Namun ada beberapa spesies yang dapat dikonsumsi. Sehingga, sulit untuk membedakan jenis jamur antara yang dapat dikonsumsi maupun tidak dikarenakan dalam satu genus jamur saja terdapat banyak perbedaan karakteristik dan ciri. Oleh sebab itu, untuk membedakan jenis jamur antara yang dapat dikonsumsi maupun tidak kami melakukan penelitian untuk melakukan klasifikasi apakah suatu jamur dapat dikonsumsi atau tidak. Pada penelitian ini, metode yang digunakan adalah dengan menggunakan metode *K-Nearest Neighbor* (KNN).

1.2 Tinjauan Pustaka

Pada penelitian sebelumnya yang digunakan sebagai acuan pada jurnal ini adalah implementasi metode klasifikasi naïve bayes untuk memprediksi besarnya penggunaan listrik rumah tangga. Dalam penelitian tersebut algoritma naïve bayes digunakan dengan memanfaatkan data training untuk mengetahui hasil probabilitas dari setiap kriteria yang ada sehingga dapat dilakukan prediksi besarnya penggunaan listrik rumah tangga. Berbeda dengan penelitian tersebut, dalam penelitian ini, kami mengimplementasikan algoritma *K-Nearest Neighbor* untuk menentukan kelayakan konsumsi jamur. Kami mengganti data yang digunakan yaitu dengan data jamur, yang terdiri dari 23 spesies jamur yang berbeda dalam family *Agaricus* dan *Lepiota*. Data ini terdiri dari 22 fitur yang mendeskripsikan bentuk dari setiap spesies jamur. Selain itu, kelas dalam data jamur tersebut, dibagi menjadi kelas jamur beracun dan tidak beracun.

Penentuan kelayakan jamur tersebut layak dikonsumsi atau tidak didasarkan pada kelasnya yaitu beracun atau tidak beracun. Menentukan kelas jamur tersebut didasarkan pada fitur-fitur yang dimiliki spesies jamur tersebut, seperti cap-shapes(bentuk tudung jamur), cap-surface (permukaan tudung jamur), cap-color (warna tudung), dan fitur lainnya. Pada penelitian ini, kami berfokus pada penerapan algoritma *K-Nearest Neighbor* untuk melakukan klasifikasi jamur yang beracun dan tidak beracun agar dapat diketahui kelayakan mengonsumsi jamur tersebut.

1.3 Landasan Teori

1.3.1 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. Secara teknis biologis, tubuh buah ini disebut basidium. Beberapa jamur aman dimakan manusia bahkan beberapa dianggap berkhasiat obat, dan beberapa yang lain beracun (Ulloa, 2001).

Jamur atau fungi dipelajari secara spesifik di dalam cabang biologi yang disebut mikologi. Para ahli mikologi mengelompokkan kingdom ini ke dalam 4 divisi. Dasar yang digunakan dalam klasifikasi ini adalah persamaan ciri-ciri. Salah satu ciri jamur adalah bereproduksi dengan spora, baik spora berflagela maupun spora tidak berflagela. Jenis-jenis jamur yang sporanya berflagela dikelompokkan dalam dunia protista yaitu myxomycotina dan Oomycotina. Sedangkan yang memiliki spora tidak berflagela dimasukkan ke dalam dunia fungi dan dibagi menjadi 3 filum, yaitu filum zygomycota, filum ascomycota, dan filum basidiomycota. Dasar klasifikasi ketiga filum tersebut adalah cara reproduksi seksual. Sedangkan jamur-jamur yang reproduksi seksualnya belum diketahui, diklasifikasikan ke dalam satu filum, yang diberi nama divisi deuteromycota. Setiap filum dibagi menjadi beberapa kelas, setiap kelas dibagi menjadi beberapa ordo dan setiap ordo dibagi menjadi beberapa famili. Contohnya adalah *Agaricus Bisporus*.

Agaricus Bisporus merupakan salah satu jenis jamur yang termasuk dalam filum basidiomycota. Dikatakan termasuk dalam filum basidiomycota dikarenakan jamur tersebut menggunakan basidium sebagai alat reproduksi untuk menghasilkan spora. Jenis jamur ini, merupakan jenis jamur yang dapat dimakan dan dapat dibudidayakan oleh masyarakat. Namun, tidak semuanya dapat dimakan dan dibudidayakan.

Agaricus Bisporus termasuk ke dalam famili *Agaricaceae* yang merupakan salah satu famili dari jamur yang dapat dibudidayakan dan dapat dikonsumsi. Namun, sebagian lain dari famili jamur ini beracun dan dapat mematikan suatu makhluk hidup. Ciri-ciri dari famili jamur ini adalah memiliki sisik pada butiran - butiran kecil di daerah tudung dengan insang yang terpisah dari batang. Namun, tidak semua genus pada famili *Agaricaceae* memiliki ciri-ciri tersebut, misalnya pada genus seperti *Lepiota* yang tidak memiliki ciri-ciri yang sama dengan familinya.

Untuk membedakan suatu jamur termasuk ke dalam kelas yang dapat dikonsumsi atau beracun adalah dengan cara membedakan jamur tersebut dari spesies jamur dengan mengenali perbedaan karakteristik antara jamur yang dapat dikonsumsi dan yang tidak dapat dikonsumsi. Ciri-ciri dari suatu spesies jamur yang tidak dapat dikonsumsi adalah jamur tersebut memiliki warna kecoklatan atau ungu kecoklatan. Namun, cara ini tidak mudah untuk dilakukan untuk membedakan suatu jamur dikarenakan terdapat jamur yang dapat dikonsumsi tetapi memiliki karakteristik seperti jamur beracun yaitu jenis jamur *Lepiota Americana*. *Lepiota Americana* merupakan jenis jamur yang dapat dikonsumsi bahkan merupakan jenis jamur yang bagus untuk dikonsumsi.

1.3.2 Data Mining

Data mining didefinisikan sebagai proses mengekstrak atau menambang pengetahuan yang dibutuhkan dari sejumlah data besar. Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti Database System, Data Warehousing, Statistical, Machine Learning, Information Retrieval, dan Komputasi Tingkat Tinggi.

Selain itu data mining didukung oleh ilmu lain seperti Neural Network, Pengenalan Pola, Spatial Data Analysis, Image Database, Signal Processing (Han dan Kamber, 2006). Untuk menghasilkan suatu model atau keputusan yang baik maka data - data yang digunakan akan diolah sedemikian rupa sesuai dengan konsep Data Mining. Tahapan dalam proses data mining adalah sebagai berikut:

1. Data Selection

Sebelum melakukan proses data mining tahapan pertama yang dilakukan adalah melakukan seleksi pada data untuk mengurangi data yang tidak ada kaitannya dengan tujuan penelitian dengan memisahkan data - data tersebut.

2. Preprocessing atau Cleaning Data

Setelah melakukan seleksi data, biasanya data hasil seleksi tidak sepenuhnya dapat langsung digunakan dikarenakan dalam data masih terdapat duplikasi data, data yang tidak konsisten atau noise pada data sehingga perlu dilakukan proses cleaning pada data.

3. Data Transformation

Setelah data dibersihkan, kemudian data ditransformasikan ke dalam bentuk yang sesuai dalam proses data mining agar data dapat digunakan sesuai dengan tujuan penelitian.

4. Data mining

Setelah melakukan transformasi data, kemudian pada tahap ini dilakukan pencarian sebuah informasi atau pola - pola untuk menentukan metode dan algoritma yang sesuai dengan tujuan penelitian.

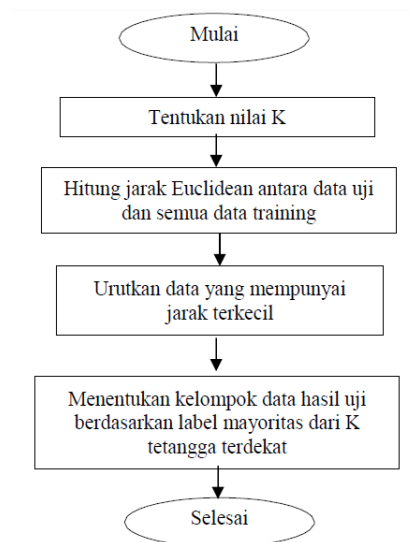
5. Interpretation atau evaluation

Setelah menentukan metode dari proses *me-mining* data, kemudian dilakukan interpretasi pada data untuk menampilkan informasi yang merupakan hasil dari proses data mining. Kemudian informasi dari proses tersebut dilakukan evaluasi dan ditampilkan dalam tampilan yang dapat dimengerti.

1.3.3 K-Nearest Neighbor (KNN)

Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga (Neighbor) terdekatnya dalam data pelatihan

Berikut urutan proses kerja KNN (Gorunescu, 2011):



1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean* (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan

$$d(x_i x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Keterangan :

$d(x_i, x_j)$: Euclidean Distance (Jarak Euclidean).

x_i, x_j : record ke-m, record ke-n

(a_r) : data ke-r

m,n : 1,2,3,...n

n : dimensi objek

3. Menentukan nilai *Nearest Neighbor*
4. Membuat klasifikasi berdasarkan *Nearest Neighbor* mayoritas

2. METODE PENELITIAN

2.1 Studi Literatur

Tahapan pertama yang dilakukan dalam penelitian ini adalah melakukan studi literatur. Studi literatur merupakan salah satu metode penelitian di mana dalam metode ini dilakukan pencarian terhadap berbagai sumber informasi yang berhubungan dengan penelitian dimulai dari mencari sumber informasi tertulis seperti jurnal, buku, majalah, maupun artikel. Studi literatur ini dilakukan untuk menentukan topik penelitian.

2.2 Sumber Data

Dalam suatu penelitian sumber data merupakan salah satu hal yang penting untuk dilakukan dikarenakan pada saat mencari sumber data, dapat ditemukan sumber informasi untuk melakukan penelitian lebih lanjut. Pengumpulan data harus dilakukan dengan baik dan benar dikarenakan apabila data yang dikumpulkan tidak sesuai maka penelitian yang akan dilakukan selanjutnya menjadi lebih sulit. Selain itu, data yang tidak sesuai dapat membuat hasil kesimpulan dari suatu penelitian yang tidak valid atau terjadi kesalahan.

Pada penelitian ini menggunakan data sekunder yang diunduh dari website UCI Machine Learning. Data yang didapatkan berisi kumpulan dataset jamur yang memiliki 22 variabel atau atribut (prediktor) dan 1 kelas atau label (klasifikasi). Data set ini memiliki 8124 record dari 23 jenis jamur yang berbeda di dalam famili Agaricus dan Lepiota yang setiap jenisnya dapat diklasifikasikan sebagai jamur yang dapat dikonsumsi (edible) dan tidak dapat dikonsumsi (poisonous).

2.3 Variabel Data

Berikut ini merupakan tabel yang berisi penjelasan terkait variabel yang terdapat di dalam dataset jamur.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Class	e = edible p = poisonous

Cap Shape	b = bell c = conical x = convex	f = flat k = knobbed s = sunken
Cap surface	f = fibrous g = grooves	y = scaly s = smooth
Cap color	n = brown b = buff c = cinnamon g = gray r = green	p = pink u = purple e = red w = white y = yellow
Bruises	t = bruises f = no	
Odor	a = almond l = anise c = creosote y = fishy f = foul	m = musty n = none p = pungent s = spicy
Gill attachment	a = attached d = descending	f = free n = notched
Gill spacing	c = close w = crowded d = distant	
Gill size	b = broad n = narrow	
Gill color	k = black n = brown b = buff h = chocolate g = gray r = green	o = orange p = pink u = purple e = red w = white y = yellow
Stalk shape	e = enlarging t = tapering	
Stalk root	b = bulbous c = club u = cup	e = equal z = rhizomorphs r = rooted
Stalk surface above ring	f = fibrous y = scaly	k = silky s = smooth
Stalk surface below ring	f = fibrous y = scaly	k = silky s = smooth
Stalk color above ring	n = brown b = buff c = cinnamon g = gray o = orange	p = pink e = red w = white y = yellow
Stalk color below ring	n = brown	p = pink

	b = buff e = red c = cinnamon w = white g = gray y = yellow o = orange
Veil type	p = partial u = universal
Veil color	n = brown w = white o = orange y = yellow
Ring number	n = none o = one t = two
Ring type	c = cobwebby n = none e = evanescent p = pendant f = flaring s = sheathing l = large z = zone
Spore print color	k = black o = orange n = brown u = purple b = buff w = white h = chocolate y = yellow r = green
Population	a = abundant s = scattered c = clustered v = several n = numerous y = solitary
Habitat	g = grasses u = urban l = leaves w = waste m = meadows d = woods p = paths

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini adalah data yang berjumlah 8.124 data jamur. Berikut adalah rincian data jamur yang belum dilakukan pengolahan atau data kotor (dirty).

Tabel 2. Data Jamur yang belum dilakukan tahap preprocessing

No	class	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	...	habitat
1	p	x	s	n	t	p	f	c	n	...	u
2	e	x	s	y	t	a	f	c	b	...	g
3	e	b	s	w	t	l	f	c	b	...	m
4	p	x	y	w	t	p	f	c	n	...	u
5	e	x	s	g	f	n	f	w	b	...	g

6	e	x	y	y	t	a	f	c	b	...	g
7	e	b	s	w	t	a	f	c	b	...	m
8	e	b	y	w	t	l	f	c	b	...	m
9	p	x	y	w	t	p	f	c	n	...	g
10	e	b	s	y	t	a	f	c	b	...	m
...
8124	e	x	s	n	f	n	a	c	?	...	l

3.1 Preprocessing Data

3.1.1 Data Cleaning

Dari data jamur yang kita dapatkan masih memiliki banyak sekali missing data (data hilang) yang ditandai dengan nilai data '?'. Maka dari itu perlu kita lakukan data cleaning dengan menerapkan metode imputasi data. Pada penelitian ini data yang digunakan bukan dalam bentuk numerik melainkan data berbentuk kategorik maka metode imputasi yang paling tepat digunakan adalah metode imputasi dengan strategi modus('most_frequent').

Tabel 3. Data Jamur yang telah dilakukan imputasi data

No	class	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	...	habitat
1	p	x	s	n	t	p	f	c	n	...	u
2	e	x	s	y	t	a	f	c	b	...	g
3	e	b	s	w	t	l	f	c	b	...	m
4	p	x	y	w	t	p	f	c	n	...	u
5	e	x	s	g	f	n	f	w	b	...	g
6	e	x	y	y	t	a	f	c	b	...	g
7	e	b	s	w	t	a	f	c	b	...	m
8	e	b	y	w	t	l	f	c	b	...	m
9	p	x	y	w	t	p	f	c	n	...	g
10	e	b	s	y	t	a	f	c	b	...	m
...
8124	e	x	s	n	f	n	a	c	b	...	l

Setelah melalui proses data cleaning, data yang sebelumnya missing atau hilang digantikan dengan nilai modus pada atribut atau kolom dimana data tersebut missing. Data yang telah dilakukan imputasi ditampilkan pada Tabel 3. Kotak kuning yang sebelumnya bernilai '?' digantikan dengan nilai modus pada variabel tersebut.

Tahapan selanjutnya adalah mengubah data yang sebelumnya bertipe kategorik menjadi numerik. Pada penelitian ini pengubahan data dilakukan berdasarkan urutan abjad, dimana abjad yang lebih awal akan memiliki nilai 0 selanjutnya 1 dan seterusnya.

Tabel 4. Data Jamur yang diubah dari kategorik menjadi numerik

No	class	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	...	habitat
1	p	5	2	4	1	6	1	0	1	...	5
2	e	5	2	9	1	0	1	0	0	...	1
3	e	0	2	8	1	3	1	0	0	...	3
4	p	5	3	8	1	6	1	0	1	...	5
5	e	5	2	3	0	5	1	1	0	...	1
6	e	5	3	9	1	0	1	0	0	...	1
7	e	0	2	8	1	0	1	0	0	...	3
8	e	0	3	8	1	3	1	0	0	...	3
9	p	5	3	8	1	6	1	0	1	...	1
10	e	0	2	9	1	0	1	0	0	...	3
...
8124	n	5	2	4	0	5	0	0	0	...	2

3.1.2 Data Reduction

Salah satu cara untuk membuat model yang baik adalah pemilihan fitur yang akan digunakan dalam pembuatan model. Dalam penelitian ini feature selection digunakan untuk melakukan reduksi data yang dianggap tidak efektif. Metode seleksi yang digunakan adalah metode Univariate Selection. Setelah dilakukan feature selection, Atribut (Prediktor) yang sebelumnya berjumlah 22 atribut menjadi 16 atribut.

Tabel 5. Fitur yang akan digunakan setelah dilakukan Data Reduction

No.	Variabel	Nilai Kontribusi
1.	Gill-Spacing	8.2679527444
2.	Odor	7.5910163218
3.	Habitat	7.5130948906
4.	Gill-Color	5.9577644688
5.	Veil-Color	5.1268256792
6.	Spore-Print-Color	3.7913272877
7.	Stalk-Shape	3.6594105493
8.	Gill-Attachment	3.5054468176
9.	Population	3.1176673588
10.	Ring-Number	2.5646335154
11.	Stalk-Surface-Above-Ring	2.2298240040
12.	Cap-Surface	2.1406854420
13.	Stalk-Surface-Below-Ring	2.0664818022
14.	Ring-Type	1.9506101464
15.	Cap-Shape	1.7508364000
16.	Gil-Size	1.6366068328

3.2 Data Transformation

Dalam pembuatan model sangat penting dilakukannya transformasi data. Nilai data pada setiap fitur mempengaruhi bagaimana model terbentuk. Setiap fitur tentunya memiliki interval nilai yang berbeda dengan fitur lainnya. Maka dari itu, transformasi data dilakukan untuk penyeragaman interval nilai dengan skala atau interval tertentu. Pada penelitian ini transformasi data dilakukan dengan metode Min-Max Normalization dengan interval nilai 0 - 1.

Tabel 6. Data jamur yang telah dilakukan transformasi data

No	class	cap-shape	cap-surface	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	habitat
1	p	1	0.666667	0.75	1	0	1	0.363636	...	0.833333
2	e	1	0.666667	0	1	0	0	0.363636	...	0.166667
3	e	0	0.666667	0.375	1	0	0	0.454545	...	0.5
4	p	1	1	0.75	1	0	1	0.454545	...	0.833333
5	e	1	0.666667	0.625	1	1	0	0.363636	...	0.166667
6	e	1	1	0	1	0	0	0.454545	...	0.166667
7	e	0	0.666667	0	1	0	0	0.181818	...	0.5
8	e	0	1	0.375	1	0	0	0.454545	...	0.5
9	p	1	1	0.75	1	0	1	0.636364	...	0.166667
10	e	0	0.666667	0	1	0	0	0.181818	...	0.5
...
8124	n	1	0.666667	0.625	0	0	0	1	...	0.333333

3.3 Pembagian Data Training dan Data Testing

Setelah data telah dilakukan preprocessing maka tahapan selanjutnya adalah pembagian data. Data dibagi menjadi data training dan testing untuk mencapai prediksi yang diinginkan. Pada penelitian ini pembagian data dilakukan dengan menggunakan metode K-Fold Cross Validation. Pada proses ini seluruh dataset akan diacak dan dibagi menjadi K bagian sama besar dengan data uji dan data latih. Metode k-fold ini menggunakan nilai $k = 10$. Pembagian data ini yang awalnya jumlah data yang digunakan sebanyak 8124 data akan dibagi menjadi 10 bagian data. Pada *fold* pertama, 9 bagian data yang berbeda akan digabungkan untuk dijadikan data training, dan satu bagian data yang tersisa akan dijadikan data testing, proses tersebut dilakukan hingga mencapai *fold* ke sepuluh.

3.4 Klasifikasi K - Nearest Neighbors

Setelah diperoleh data yang digunakan untuk pengujian kemudian dicari klasifikasi dari masing-masing data testing. Karena banyaknya data yang akan digunakan maka untuk penghitungan jarak euclidean akan diberikan contoh penghitungannya. Sebagai contoh penghitungan jarak euclidean, data

yang kita gunakan, yaitu 10 data jamur sebagai data training dan 1 data jamur sebagai data testing. Contoh pembagian data adalah seperti pada Tabel 7.

Tabel 7. Contoh Sample Data Training dan Data Testing

No	class	cap- shape	cap- surface	odor	gill- attachment	gill- spacing	gill- size	gill- color	...	habitat
<i>Data training</i>										
A1	p	1	0.666667	0.75	1	0	1	0.363636	...	0.833333
A2	e	1	0.666667	0	1	0	0	0.363636	...	0.166667
A3	e	0	0.666667	0.375	1	0	0	0.454545	...	0.5
A4	p	1	1	0.75	1	0	1	0.454545	...	0.833333
A5	e	1	0.666667	0.625	1	1	0	0.363636	...	0.166667
A6	e	1	1	0	1	0	0	0.454545	...	0.166667
A7	e	0	0.666667	0	1	0	0	0.181818	...	0.5
A8	e	0	1	0.375	1	0	0	0.454545	...	0.5
A9	p	1	1	0.75	1	0	1	0.636364	...	0.166667
A10	e	0	0.666667	0	1	0	0	0.181818	...	0.5
<i>Data Testing</i>										
A8124	X	1	0.666667	0.625	0	0	0	1	...	0.333333

Selanjutnya kita akan menerapkan K-NN sesuai dengan tahapannya :

1. Menetapkan parameter K (jumlah tetangga terdekat) yang akan digunakan. Pada penelitian ini, untuk mendapatkan model yang optimal maka parameter K yang akan digunakan yaitu K = 3, 5, 7, 9, 11.
2. Kemudian menghitung jarak euclidean (euclidean distance) dari setiap objek training yaitu A1-A10 terhadap data testing yang telah ditentukan, yaitu A8124. Rumus yang akan digunakan terdapat pada sub bab 1.3.3 poin yang kedua.

$$\begin{aligned}
 d = & \sqrt{(1 - 1)^2 + (0,666667 - 0,666667)^2 + (0,75 - 0,625)^2 + (1 - 0)^2 + (0 - 0)^2 + (1 - 0)^2} \\
 & + (0,363636 - 1)^2 + (0 - 0)^2 + (0,666667 - 0,666667)^2 \\
 & + (0,666667 - 0,666667)^2 + (0,666667 - 0,333333)^2 + (0,5 - 0,5)^2 + (1 - 1)^2 \\
 & + (0,25 - 0,5)^2 + (0,6 - 0,2)^2 + (0,833333 - 0,333333)^2 = 1,733262
 \end{aligned}$$

Jadi, jarak Euclidean yang diperoleh dari contoh perhitungan adalah 1,733262. Hasil perhitungan jarak dari sepuluh data training diatas terhadap data testing ditampilkan pada Tabel 8.

Tabel 8. Hasil perhitungan jarak A8124 terhadap A1-A10

No	class	cap- shape	cap- surface	odor	gill- attachment	gill- spacing	gill- size	gill- color	...	Distance
A1	p	1	0.666667	0.75	1	0	1	0.363636	...	1.733261578
A2	e	1	0.666667	0	1	0	0	0.363636	...	1.410708422
A3	e	0	0.666667	0.375	1	0	0	0.454545	...	1.598291463
A4	p	1	1	0.75	1	0	1	0.454545	...	1.73432079
A5	e	1	0.666667	0.625	1	1	0	0.363636	...	2.144638257
A6	e	1	1	0	1	0	0	0.454545	...	1.428511868
A7	e	0	0.666667	0	1	0	0	0.181818	...	1.816985479
A8	e	0	1	0.375	1	0	0	0.454545	...	1.669025611
A9	p	1	1	0.75	1	0	1	0.636364	...	1.679391598
A10	e	0	0.666667	0	1	0	0	0.181818	...	1.849712472

Setelah didapatkan hasil perhitungan jarak untuk semua data training, kemudian untuk mendapatkan K tetangga terdekat maka diurutkan objek – objek (A1 – A10) mulai dari jarak terkecil ke terbesar seperti pada Tabel 9.

Tabel 9. Hasil perhitungan yang telah diurutkan

No	class	cap- shape	cap- surface	odor	gill- attachment	gill- spacing	gill- size	gill- color	...	Distance
A2	e	1	0.666667	0	1	0	0	0.363636	...	1.410708
A6	e	1	1	0	1	0	0	0.454545	...	1.428512
A3	e	0	0.666667	0.375	1	0	0	0.454545	...	1.598291
A8	e	0	1	0.375	1	0	0	0.454545	...	1.669026
A9	p	1	1	0.75	1	0	1	0.636364	...	1.679392
A1	p	1	0.666667	0.75	1	0	1	0.363636	...	1.733262
A4	p	1	1	0.75	1	0	1	0.454545	...	1.734321
A7	e	0	0.666667	0	1	0	0	0.181818	...	1.816985
A10	e	0	0.666667	0	1	0	0	0.181818	...	1.849712
A5	e	1	0.666667	0.625	1	1	0	0.363636	...	2.144638

Setelah diperoleh hasil perhitungan, langkah selanjutnya menentukan hasil klasifikasi berdasarkan nilai K yang telah kita tentukan sebelumnya, yaitu : K=3, K=5, K=7, K=9 dan K=11. Sebagai contoh, jika kita menggunakan nilai K=3. Maka, tiga data terdekat yaitu tiga data dengan jarak (distance) euclidean terkecil akan digunakan untuk proses klasifikasi pada data testing A8124. Ketiga data training dengan jarak terkecil adalah A2, A6 dan A3. Langkah selanjutnya menentukan kelas mayoritas dari ketiga kelas data tersebut. Dimana kelas A2 = e, kelas A6 = e dan kelas A3 = e. Maka mayoritas dari ketiga kelas tersebut adalah e (edible). Jadi, data testing jamur A8124 dapat diklasifikasikan sebagai jamur yang dapat dikonsumsi.

Selanjutnya untuk menentukan keakuratan model yang telah dibentuk, maka langkah yang akan dilakukan adalah menentukan akurasi dari model. Tingkat akurasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Dalam penelitian ini, perhitungan akurasi dilakukan dengan menggunakan evaluasi model Confusion Matrix. Tingkat akurasi Confusion Matrix didapatkan dengan membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Pada Confusion Matrix terdapat 4 istilah representasi hasil proses klasifikasi, yaitu : True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Hasil dari akurasi, sensitivity dan specificity pada model yang telah dibentuk pada penelitian ini ditampilkan pada Tabel 10.

Tabel 10. Hasil akurasi berdasarkan parameter K

K (Jumlah tetangga)	Akurasi	Sensitivity	Specificity
3	99.0148	100	98.0998
5	97.6601	100	95.4869
7	97.6601	100	95.4869
9	97.1675	99.4885	95.0119
11	97.0443	99.2327	95.0119

Berdasarkan Tabel 10 hasil akurasi terbaik yaitu menggunakan parameter K=3. Model terbaik yang dibentuk untuk klasifikasi jamur pada penelitian ini mendapatkan tingkat akurasi mencapai 99,01%.

4. KESIMPULAN

Setelah dilakukan analisis, perancangan, pengujian serta implementasi yang telah dilakukan pada dataset jamur maka didapatkan kesimpulan terhadap penentuan suatu jamur dapat dikonsumsi atau tidak, sebagai berikut :

1. Dihasilkan 16 variabel atau 16 ciri dari suatu jamur yang memiliki kontribusi besar untuk menentukan suatu jamur dapat dikonsumsi atau tidak.
2. Dengan menerapkan perbandingan model yang menggunakan K=3, K=5, K=7, K=9 dan K=11 penelitian mendapatkan model yang optimal.
3. Hasil klasifikasi terhadap dataset jamur dengan melakukan perhitungan menggunakan algoritma KNN diperoleh hasil klasifikasi dengan nilai akurasi sebesar 99.01% yaitu dengan nilai parameter K = 3. Sehingga hasil klasifikasi dengan algoritma KNN ini dapat dikatakan sangat baik untuk menentukan kelayakan suatu jamur apakah dapat dikonsumsi (edible) atau tidak dapat dikonsumsi (poisonous).

5. SARAN

Setelah melihat hasil dari penelitian yang dilakukan ini, maka penulis memberikan saran yang diharapkan dapat menjadi bahan pertimbangan untuk melakukan penelitian selanjutnya. Berikut merupakan saran-saran tersebut :

1. Dapat menggunakan algoritma yang lainnya, sehingga akan mendapatkan perbandingan nilai akurasi yang lebih baik dari penelitian sebelumnya.

2. Dapat menggunakan cakupan data yang lebih luas sehingga pengelompokkan jamur yang dapat dikonsumsi dan tidak dapat dikonsumsi lebih akurat.

DAFTAR PUSTAKA

<https://www.kaggle.com/uciml/mushroom-classification>

<https://id.wikipedia.org/wiki/Agaricus>

Putra, Bayu Mahardika. 2008. *Klasifikasi Jamur ke Dalam Kelas Dapat Dikonsumsi atau Beracun Menggunakan Algoritma VF15 (Studi Kasus : Famili Agaricus dan Lepiota)*. Repository IPB.

<https://ejournal.undiksha.ac.id/index.php/JST/article/download/8549/5789>

Ulloa M, Richard TH. 2001. *Illustrated Dictionary of Mycology*. Minnesota: APS Press.

Han, Jiawei dan Kamber, Micheline. 2006. *Data Mining : Concept and Techniques Second Edition*. Morgan Kaufmann Publishers