Data Mining Pada Klasifikasi Jamur Menggunakan Algoritma C.45 Berdasarkan Karakteristik Morfologi Mushroom

Rofi Fathurrohman[1], Miftahul Fajri S[2], Salman M Cahyono[3], Dzulqa Fauzan Abdillah[4], Vinan Ramadhan[5]
[1,2,3,4,5]Program Studi Sistem Informasi, Fakultas Teknik Infomatika
Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98 Senen, Jakarta Pusat-10420, Indonesia

e-mail: [1]19236037@bsi.ac.id, [2]19236040@bsi.ac.id, [3]19236045@bsi.ac.id, [4]19236043@bsi.ac.id, [5]19236042@bsi.ac.id

Artikel Info : Diterima : 15-12-2023 | Direvisi : 04-03-2024 | Disetujui : 28-06-2024

Abstrak - Jamur merupakan salah satu bahan pangan yang populer di Indonesia, namun tidak semua jamur dapat dikonsumsi karena beberapa jenis jamur bersifat racun. Oleh karena itu, diperlukan metode untuk mengklasifikasikan jamur menjadi dapat dimakan dan beracun. Penelitian ini bertujuan untuk melakukan klasifikasi jamur menggunakan algoritma klasifikasi data mining yaitu algoritma C.45 berdasarkan karakteristik morfologi jamur. Data jamur dikumpulkan dari berbagai sumber dan memiliki 22 atribut, termasuk bau, warna cetakan spora, akar tangkai, ukuran insang, dan jarak insang. Algoritma C.45 digunakan untuk membangun pohon keputusan yang dapat mengklasifikasikan jamur. Hasil penelitian menunjukkan bahwa algoritma C.45 memiliki akurasi sebesar 99.90%. Interpretasi pohon keputusan menunjukkan bahwa jamur yang memiliki : berbau amis, apak, busuk, creosof, menyengat, pedas, dan tidak berbau, warna cetakan spora hijau dan putih, akar tangkai klub dan bulat, ukuran insang sempit dan jarak insang dekat adalah jamur beracun. Sementara itu, jamur yang tidak memiliki karakteristik tersebut adalah jamur yang dapat dimakan. Kesimpulannya, algoritma C.45 dapat digunakan sebagai metode efektif untuk mengklasifikasikan jamur dan membantu masyarakat dalam membedakan jamur yang dapat dimakan dan beracun.

Kata Kunci : Algoritma C4.5, Jamur, Klasifikasi

Abstracts *- Mushrooms are one of the popular food ingredients in Indonesia, but not all mushrooms can be consumed because some types of mushrooms are toxic. Therefore, a method is needed to classify mushrooms into edible and toxic. This research aims to classify mushrooms using a data mining classification algorithm, namely the C.45 algorithm based on the morphological characteristics of mushrooms. The mushroom data were collected from various sources and had 22 attributes, including odor, spore mold color, stalk root, gill size, and gill spacing. The C.45 algorithm was used to build a decision tree that can classify mushrooms. The results show that the C.45 algorithm has an accuracy of 99.90%. The interpretation of the decision tree shows that mushrooms that have: fishy, musty, rotten, creosophical, pungent, spicy, and odorless, green and white spore mold colors, club and round stalk roots, narrow gill size and close gill spacing are poisonous mushrooms. Meanwhile, mushrooms that do not have these characteristics are edible mushrooms. In conclusion, the C.45 algorithm can be used as an effective method to classify mushrooms and help people in distinguishing edible and poisonous mushrooms*

Keywords : C4.5 Algoritm, Classification, Mushroom

PENDAHULUAN

Jamur adalah makhluk yang bersifat *eukariotik heterotrofik* uniseluler dan multiseluler yang membutuhkan molekul organik untuk nutrisinya (Reece et al., 2014). Dengan berbagai jenis jamur liar tumbuh di hutan dan padang rumput, dan umumnya dikonsumsi oleh penduduk setempat (Lancet, 1980). Diperkirakan terdapat 1,5 juta spesies jamur di dunia ini, namun hanya 70.000 yang telah ditemukan dan telah diidentifikasi, dan sekitar 1,43 juta (95%) belum teridentifikasi (Hawksworth & Rossman, 1997).

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Jamur termasuk ke dalam kategori makanan bernilai gizi tinggi serta rendah kalori yang mengandung protein, vitamin, dan mineral berkualitas tinggi. Hal ini menjadikan jamur sebagai sumber alami Hal ini menjadikan jamur sebagai sumber alami yang berharga dan memiliki potensi untuk dimanfaatkan sebagai makanan maupun obat-obatan. Masyarakat tradisional mengenali penggunaan obat dari jamur yang dapat dimakan dan jamur liar, yang sekarang sedang diuji bioaktivitasnya untuk berbagai penyakit. Khasiat penyembuhan melalui pemanfaatan jamur digunakan dalam pengobatan alternatif. Jamur yang dapat dimakan telah dianggap sebagai makanan yang baik untuk penderita obesitas dan penderita diabetes untuk mengurangi *hiperglikemia* karena kandungan seratnya yang tinggi, rendah lemak, dan rendah pati. Jamur ini juga dianggap memiliki efek antioksidan, *kardiovaskular, hiperkolesterol,* antibakteri, *hepatoprotektif,* dan antikanker (Khatun et al., 2012).

Metode identifikasi jamur dapat dibagi menjadi beberapa jenis, yaitu metode morfologi, metode mikroskopis, dan metode biokimia. Ciri-ciri morfologi yang dimaksud adalah bentuk payung, warna, tekstur, dan aspek-aspek lain yang dapat diamati (Zubair & Muslikh, 2017). Jenis jamur yang termasuk dalam genus *agaricus* dan *lepiota* adalah jamur yang banyak ditemukan di alam, bahkan jamur jenis ini banyak dibudidayakan, atau dikonsumsi. Spesies *agaricus* dan *lepiota* dapat hidup di tempat terbuka, tumbuh dalam berbagai warna, bentuk, tekstur, dan ukuran, serta ada yang beracun dan ada pula yang tidak memiliki racun (Halling et al., 2004).

Pada pengujian sebelumnya, seorang peneliti harus mengumpulkan data dari ciri-ciri jamur secara teliti dan terus menerus. Jika data tersebut dikumpulkan secara rutin, maka data yang terkumpul akan menjadi sangat banyak, bahkan jumlah data yang terkumpul dari hasil pengamatan dan uji coba bisa mencapai hingga ribuan (Ihsan & Yusuf, 2013).

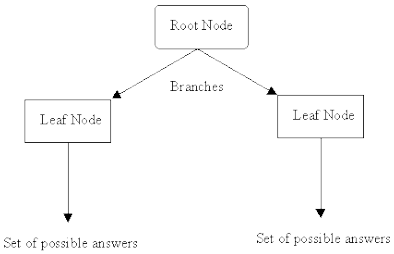
Masalah yang timbul adalah bahwa beberapa jamur berbahaya memiliki ciri-ciri morfologi yang sangat mirip dengan jamur yang dapat dikonsumsi. Oleh karena itu, mengidentifikasi jamur ini, apakah beracun atau dapat dikonsumsi, akan sulit jika hanya dilihat dengan mata telanjang. Mengenali jamur yang boleh dikonsumsi sangatlah penting, karena jika jamur yang teridentifikasi beracun dapat menyebabkan keracunan dan bisa mengganggu kerja organ pencernaan jika dikonsumsi (Haksoro & Setiawan, 2022).

Untuk mencegah dan menghindari risiko tersebut, sangat penting untuk mengklasifikasikan jamur dengan akurat untuk membantu menentukan jamur mana yang dapat dimakan dan mana jamur yang beracun. Ada banyak metode untuk menentukan jamur mana yang dapat dimakan dan mana yang berbahaya. Salah satu fitur penting yang digunakan sebagai standar dalam mengenali jamur adalah penggunaan data mining (Rudiyanto et al., 2023).

Praktik pengumpulan informasi dan pola yang dapat digunakan dan dimanfaatkan dari data dalam jumlah besar dikenal sebagai data mining. Data mining mencakup pengumpulan, ekstraksi, analisis, dan statistik data. Penemuan pengetahuan *(Knowledge discovery)*, ekstraksi pengetahuan *(Knowledge extraction)*, analisis data/pola *(Data/pattern analysis)*, pengumpulan informasi *(Information harvesting)*, dan istilah lainnya digunakan untuk menggambarkan data mining (Arhami et al., 2020). Adapun metode umum data mining meliputi deskripsi, prediksi, estimasi, klasifikasi, *clustering* (pengelompokan), dan asosiasi. Prediksi dan klasifikasi adalah metode yang banyak digunakan dalam data mining untuk memeriksa data yang dapat mengkarakterisasi kelas data atau meramalkan data di masa depan (Ucha Putri et al., 2021).

Fokus utama untuk analisis data mining dalam konteks penelitian ini adalah klasifikasi jamur. Klasifikasi adalah salah satu pendekatan analisis data mining yang dapat digunakan untuk mengelompokkan data ke dalam kelas atau kelompok tertentu. Dalam konteks penelitian ini, kelas yang dimaksud adalah jamur yang dapat dikonsumsi dan jamur yang beracun. Hal ini dikenal sebagai pembelajaran yang diawasi karena kelas dari setiap observasi dipilih terlebih dahulu sebelum kategorisasi (Saputri, n.d.). Data morfologi jamur dapat dipetakan ke dalam kelas atau kelompok tertentu dengan menggunakan algoritma klasifikasi data mining. Algoritma klasifikasi data mining yang digunakan dalam penelitian ini adalah Algoritma C4.5. Algoritma C4.5 merupakan algoritma decision tree yang dikembangkan oleh Ross Quinlan. Algoritma ini dipilih karena memiliki beberapa keunggulan, yaitu: kemampuan untuk menangani data yang tidak lengkap dan data yang tidak seimbang dari data *Mushrooms Classification Dataset*.

Menggunakan decision tree dalam penelitian ini memungkinkan teknik prediksi dan klasifikasi yang kuat yang dapat membagi kumpulan data yang besar menjadi himpunan catatan yang lebih kecil dengan menggunakan serangkaian aturan keputusan (Fayyad et al., 1996).



Sumber: Penelitian (2023)

Gambar 1. *Decision Tree*

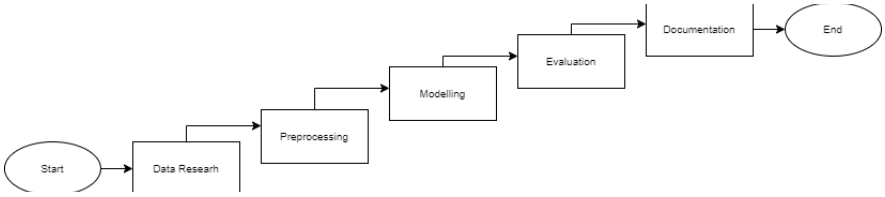
a. *Root node* (akar): tujuan akhir atau keputusan besar yang ingin diambil.
b. *Branches* (ranting): berbagai pilihan tindakan.
c. *Leaf node* (daun): kemungkinan hasil atas setiap tindakan.

Menurut penelitian sebelumnya yang dilakukan oleh Ortega *et al*. (2020) mendapatkan nilai akurasi yang dilakukan dari data tentang jamur dengan metode komparasi *(Comparison Techniques)* dari beberapa model yang digunakan dan ketepatan akurasi antara lain 87,8% Logistic Regretion, 86,5% Naïve Bayes, 88,2% Decion Tree dan 87,9% K-Nearest Neighbor (KNN) (Ortega et al., 2020), menetapkan hasil akurasi paling besar dengan menggunakan Decision tree.

Hasil akurasi lain melalui penelitian oleh Wati, Fauzan, dan Harliana (2022), metode klasifikasi dilakukan dan memaparkan hasil akurasi bahwa dengan menggunakan metode Decision Tree (DT) C5.0 pada skenario 3 menghasilkan akurasi sebesar 97,11%, sedangkan dengan menggunakan algoritma DT C4.5 menghasilkan akurasi sebesar 97,05%. Hal ini menunjukkan bahwa algoritma DT C5.0 untuk mengidentifikasi jamur berbahaya lebih akurat (Wati et al., 2022). Selanjutnya hasil akurasi yang diperoleh dari penelitian oleh Nor Sarizan F, (2021) mengusulkan perolehan hasil akurasi menggunakan Decision tree untuk prediksi jamur yang dapat dimakan atau beracun mencapai akurasi 99,81% (Nor Sarizan et al., 2021).

METODE PENELITIAN

Sebagai suatu rangkaian proses tahapan pertama yang berperan penting dalam proses perencanaan penelitian adalah menggambarkan diagram, guna untuk mempermudah serta dapat menentukan setiap langkah-langkah proses penelitian.



Sumber: Penelitian (2023)

Gambar 2. *Flowchart* Metode Penelitian

Sebagai suatu rangkaian proses tahapan pertama yang berperan penting dalam proses perencanaan penelitian adalah menggambarkan diagram, guna untuk mempermudah serta dapat menentukan setiap langkah-langkah proses penelitian.

1. Eksplorasi Data

Sebagai suatu rangkaian proses tahapan pertama yang berperan penting dalam proses perencanaan penelitian adalah menggambarkan diagram, guna untuk mempermudah serta dapat menentukan setiap langkah-langkah proses penelitian.

Penelitian ini mengacu pada studi literatur atau kepustakaan yang dilakukan dengan cara menelusuri dan mengevaluasi literatur guna mencari perkembangan penelitian dan metode penelitian yang telah digunakan sebelumnya untuk mendapatkan orientasi yang ada dalam permasalahan. Selanjutnya adalah melakukan penentuan objek yang akan diteliti untuk mengklasifikasi karakteristik *morfologi* jamur *(mushroom).* Pada tahap ini akan menjadi langkah awal yang mengharuskan peneliti untuk menentukan objek penelitian sebagai ruang lingkup metodelogi penelitian dengan eksplorasi data *(Data Research).* Tahap ini akan menghasilkan model

yang digunakan pada penelitian sebelumnya untuk menghasilkan tujuan model yang lebih baik.

Dataset yang dieksplorasi dan dijadikan bahan sebagai data olah pada penelitian ini bersumber dari *UCI Machine Learning*, yang diakses pada bulan November 2023. Berisikan 23 atribut yang meliputi karakteristik *morfologi* dan termasuk atribut target *(Class)* yang disajikan dalam tabel 1 Informasi atribut Dataset *Mushroom*.

Tabel 1. Informasi atribut

No	Atribut	*Value*
1.	*Cap shape*	Cekung, Cembung, Datar, Kerucut, Lonceng, menonjol
2.	*Cap surface*	Alur, Berserat , Bersisik, Halus
3.	*Cap color*	Abu-abu, *Buff*, Coklat, Hijau, Kayu Manis, Kuning, Merah, Merah jambu, Putih, Ungu
4.	*Bruises*	Memar, Tidak
5.	*Odor*	Adas, Manis, *Almond*, Amis, Apak, Busuk, *Creosot*, Menyengat, Pedas, Tidak ada
6.	*Gill attachment*	Terpasang, Menurun, Bebas, Berlekuk
7.	*Gill spacing*	Dekat, Padat, Jauh
8.	*Gill size*	Lebar, Sempit
9.	*Gill color*	Abu-abu, *Buff*, *Chocolate*, Coklat, Hijau, Hitam, Kuning, Merah, Merah Jambu, Oren, Putih, Ungu
10.	*Stalk shape*	Membesar, Meruncing
11.	*Stalk root*	Berakar, Blok, Bulat, Sama, Cawan, Rimpang, *Missing*
12.	*Stalk surface above ring*	Berserat, Bersisik, Halus, Lembut
13.	*Stalk surface below ring*	Berserat, Bersisik, Halus, Lembut
14.	*Stalk color above ring*	Abu-abu, *Buff*, Coklat, Kayu manis, Kuning, Merah, Merah Jambu, Oren, Putih
15.	*Stalk color below ring*	Abu-abu, *Buff*, Coklat, Kayu manis, Kuning, Merah, Merah Jambu, Oren, Putih
16.	*Veil type*	Sebagian, Universal
17.	*Veil color*	Coklat, Kuning, Oren, Putih
18.	*Ring number*	Dua, Satu, Tidak ada
19.	*Ring type*	Besar, *Evanescent*, *Flaring*, *Liotin*, Tidak ada, Sarang laba-laba, Selubung, zona
20.	*Spore print color*	*Buff*, Coklat, *Chocolate*, Hijau, Hitam, Kuning, Oren, Putih, Ungu
21.	*Population*	Banyak, Beberapa, Berkelompok, Berlimpah, Tersebar, Tunggal
22.	*Habitat*	Daun, Jalan, Kayu, Padang, Rumput, Perkotaan, Rumput, Sampah
23.	*Class*	Dapat dimakan, Beracun

Sumber: (UCI Machine Learning Repository, 1981)

2. *Preprocessing*

Sebelum memulai analisis, hal utama yang perlu dilakukan adalah mengevaluasi data, yang biasanya disebut sebagai tahap *preprocessing*.

a. *Feature Selection*

Dari pengetahuan sebelumnya, bahwa terdapat banyak serangkaian atribut dari *Mushroom* dataset, yang memungkinkan adanya atribut yang tidak berdampak, sehingga ini akan menjadi ketidakefektifan pada tahap analisis. Oleh karena itu dilakukannya *feature selection* atau pemilihan variabel menjadikannya sangat penting pada analisis data mining.

Penentuan atribut dilakukan dengan menggunakan *Information Gain* dengan menghitung nilai semua atribut. *Information gain* adalah metode seleksi fitur yang menghitung penurunan *entropy* dari sebelum dan sesudah pemisahan. Hal ini digunakan untuk menentukan fitur mana yang akan digunakan dan mana yang akan dibuang (Tangirala, 2020).

Seleksi fitur dengan akuisisi informasi terjadi dalam tiga tahap:

a) Tentukan nilai *information gain* untuk setiap atribut pada dataset asli.
b) Tentukan *cutoff* yang diperlukan. Hal ini memungkinkan karakteristik dengan bobot yang sama atau lebih besar dari ambang batas untuk dipertahankan sementara membuang atribut dengan bobot kurang dari ambang batas.
c) Pengurangan atribut akan memperbaiki dataset.

Dari Persamaan 1 digunakan untuk menghitung nilai *Information Gain*:

$$Info\ (D) = -\sum_{i=1}^{m} pi\ log_2\ (pi)$$

Dimana *D* adalah himpunan kasus, *M* adalah jumlah partisi *D*, dan *pi* adalah probabilitas dari sebuah *tuple D*.

Sementara itu, Persamaan 2 digunakan untuk menghitung nilai entropi setelah pemisahan.

$$info_A = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} X\ I\ (D_j)$$

b. *Imputation*

Setelah tahap pemilihan fitur, tahap selanjutnya yang akan dilakukan adalah imputasi data, tahap ini dilakukan untuk menghitung jumlah *record* dari setiap atribut dan ini adalah tahap identifikasi. Tujuan dari evaluasi ini adalah untuk mengetahui apakah ada *missing value* pada suatu atribut. Hasil pemeriksaan missing value pada dataset *mushroom* ditunjukkan pada Tabel 2 Pendeteksian jumlah *missing value*.

Tabel 2. Pendeteksian jumlah *missing value*

No	Variabel	Jumlah m*issing value*
1.	*Cap shape*	0
2.	*Cap surface*	0
3.	*Cap color*	0
4.	*Bruises*	0
5.	*Odor*	0
6.	*Gill attachment*	0
7.	*Gill spacing*	0
8.	*Gill size*	0
9.	*Gill color*	0
10.	*Stalk shape*	0
11.	*Stalk root*	2480
12.	*Stalk surface above ring*	0
13.	*Stalk surface below ring*	0
14.	*Stalk color above ring*	0
15.	*Stalk color below ring*	0
16.	*Veil type*	0
17.	*Veil color*	0
18.	*Ring number*	0
19.	*Ring type*	0
20.	*Spore print color*	0
21.	*Population*	0
22.	*Habitat*	0
23.	*Class*	0

Sumber: Penelitian (2023)

Pada dataset yang telah disajikan, terhitung jumlah *record* yang memiliki *missing value* dan teridentifikasi yaitu terdapat 2480 *record* yang menunjukkan bahwa variabel akar tangkai *(stalk root)* terdapat *missing value*. Oleh karena itu, dilakukan imputasi data *(data imputation) missing value* dengan modus, Berdasarkan tabel diatas, dalam hal ini adalah *missing value* akan akan dilakukan transformasi nilai berdasarkan nilai yang paling banyak muncul dalam suatu variabel dan didasarkan pada jenis klasifikasi *(Class)*. Ini dilakukan karena dari data set *mushroom* bersifat kategorik.

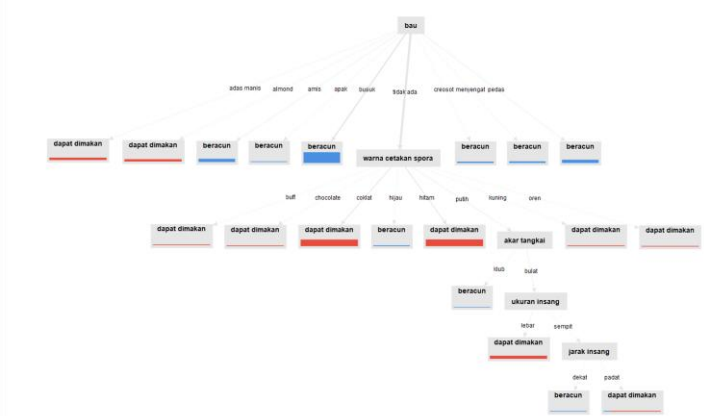
3. Pemodelan *(Modeling)*

Setelah tahapan penyiapan data dan identifikasi *missing value* pada tahap *preprocessing* telah dilalui, selanjutnya dilakukan representasi yang terstruktur dari serangkaian keputusan yang diambil berdasarkan atribut dan kelas.

Praktik evaluasi pada penelitian dilakukan menggunakan *Machine Learning* RapidMiner Studio 10.2, dimana dataset yang sudah diproses sebelumnya di-import ke dalam *machine learning*. Evaluasi dilakukan dengan mengimplementasikan metode *decision tree* yang dihasilkan melalui mekanisme penerapan dari algoritma C4.5, dengan kata lain algoritma C4.5 digunakan untuk menghasilkan *decision tree*.

Umumnya algoritma C4.5 digunakan untuk menghasilkan output berupa decision tree, Decision Tree adalah struktur pohon di mana setiap node mewakili atribut yang diuji, setiap cabang menunjukkan pembagian hasil pengujian, dan node daun mewakili kelompok kelas tertentu. Node akar adalah node tingkat teratas dari Decision Tree, dan biasanya merupakan atribut yang memiliki pengaruh paling besar pada kelas tertentu. Nilai-nilai atribut akan diperiksa selama proses klasifikasi dengan mempelajari jalur dari node akar ke node daun dan kemudian memprediksi kelas yang dimiliki oleh data baru yang spesifik (Rosela, 2019).

Gambar 3 Grapik *Decision tree* yang merupakan hasil evaluasi penerapan *Decision tree* yang diperoleh dan diproses menggunakan software RapidMiner.



Sumber: Penelitian (2023)

Gambar 3. Grapik *decision tree*

Berikut adalah beberapa ciri morfologi jamur beracun yang perlu diperhatikan: berbau amis, apak, busuk, *creosof*, menyengat, pedas, dan tidak berbau, warna cetakan spora hijau dan putih, akar tangkai klub dan bulat, ukuran insang sempit dan jarak insang dekat. Bau *(odor)* berada diposisi simpul akar *(root node)* sebagai karakteristik *morfologi* yang paling mendominasi dan menjadi kebergantungan oleh karakteristik lainnya sehingga bisa membedakan jamur yang dapat dimakan dan jamur yang beracun, begitupun berdasarkan deskripsi dari *decision tree* disajikan dalam Gambar 4 Deskripsi.

Tree

```
bau = adas manis: dapat dimakan {beracun=0, dapat dimakan=400}
bau = almond: dapat dimakan {beracun=0, dapat dimakan=400}
bau = amis: beracun {beracun=576, dapat dimakan=0}
bau = apak: beracun {beracun=36, dapat dimakan=0}
bau = busuk: beracun {beracun=2160, dapat dimakan=0}
bau = creosot: beracun {beracun=192, dapat dimakan=0}
bau = menyengat: beracun {beracun=256, dapat dimakan=0}
bau = pedas: beracun {beracun=576, dapat dimakan=0}
bau = tidak ada
|   warna cetakan spora = buff: dapat dimakan {beracun=0, dapat dimakan=48}
|   warna cetakan spora = chocolate: dapat dimakan {beracun=0, dapat dimakan=48}
|   warna cetakan spora = coklat: dapat dimakan {beracun=0, dapat dimakan=1344}
|   warna cetakan spora = hijau: beracun {beracun=72, dapat dimakan=0}
|   warna cetakan spora = hitam: dapat dimakan {beracun=0, dapat dimakan=1296}
|   warna cetakan spora = kuning: dapat dimakan {beracun=0, dapat dimakan=47}
|   warna cetakan spora = oren: dapat dimakan {beracun=0, dapat dimakan=48}
|   warna cetakan spora = putih
|   |   akar tangkai = bulat
|   |   |   ukuran insang = lebar: dapat dimakan {beracun=0, dapat dimakan=529}
|   |   |   ukuran insang = sempit
|   |   |   |   jarak insang = dekat: beracun {beracun=32, dapat dimakan=0}
|   |   |   |   jarak insang = padat: dapat dimakan {beracun=8, dapat dimakan=48}
|   |   akar tangkai = klub: beracun {beracun=8, dapat dimakan=0}
```

Sumber: Penelitian (2023)

Gambar 4. Deskripsi

4. Evaluasi

Langkah penting diambil untuk menentukan keakuratan model *decision tree* dalam membedakan jamur yang dapat dimakan dan jamur yang beracun untuk tujuan melakukan pengujian sistem dengan langkah evaluasi sistem yang dapat dijangkau dan dianalisis. Untuk melakukan langkah tersebut, peneliti menggunakan *confusion matrix* yang dibuat setelah metode *cross-validation* yang membagi seluruh kumpulan data ke dalam jumlah bagian yang diperlukan. Metode ini digunakan untuk mencegah kelebihan dan kekurangan *fitting*. Jumlah potongan menunjukkan jumlah angka atau iterasi pengujian klasifikator yang harus dilakukan. Satu *fold* akan berfungsi sebagai set tes, atau data tes, dan *fold* lainnya akan berfungsi sebagai set pelatihan. Sampai jumlah *fold* yang ditetapkan, prosedur ini diulangi.

Confusion Matrix adalah alat untuk mengukur ketepatan konsep data mining (Mayadewi & Rosely, 2015). Tabel 3 *Confusion Matrix* menunjukkan hasil validasi.

Tabel 3. *Confusion Matrix*

	False Positive (FP)	*False Negative* (FN)
True Positive (TP)	TP	FN
True Negative (TN)	FP	TN

Sumber: Penelitian (2023)

Confusion matrix dapat menampilkan jumlah jamur beracun yang terklasifikasi dengan benar dari hasil proses sistem (TP), jamur yang dapat dimakan yang terklasifikasi dengan benar dari hasil proses sistem (TN), jamur beracun yang terklasifikasi dengan benar dari hasil proses sistem (FP), dan jamur beracun yang klasifikasi dengan benar dari hasil proses sistem (FN).

Dengan kata lain, angka akurasi adalah perbandingan antara data yang diklasifikasikan dengan benar dan semua data. Ketepatan proporsi yang tepat dapat digunakan untuk membuat prediksi. Akurasi, *recall*, presisi, dan skor F1 adalah ukuran evaluasi kinerja yang paling umum digunakan dalam literatur. Kriteria-kriteria ini digunakan dalam penelitian ini untuk menilai kinerja dan keberhasilan model yang dihasilkan. Persamaan 5-7 digunakan untuk menghitung ukuran-ukuran ini.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} X\ 100\% \tag{1}$$

$$Presisi = \frac{TP}{TP + FP} X\ 100\% \tag{2}$$

$$Recall = \frac{TP}{TP + FN} X\ 100\% \tag{3}$$

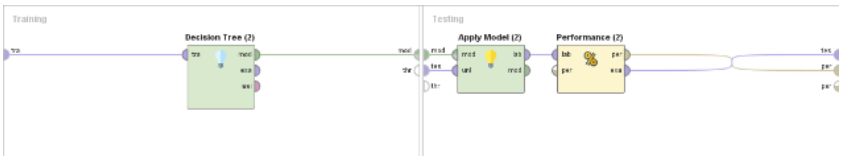
$$F1\ Score = 2x\ \frac{Presisi\ x\ Recall}{Presisi\ +\ Recall} X\ 100\% \tag{4}$$

HASIL DAN PEMBAHASAN

Setelah menyelesaikan langkah-langkah pra-pemrosesan data, pemilihan fitur, dan pemodelan, langkah berikutnya adalah menerapkan evaluasi akhir dari hasil klasifikasi untuk menentukan kategori jamur yang dapat dimakan dan jamur yang beracun. Ini menghasilkan nilai akurasi *(accuracy)*, presisi *(precision)*, dan *recall* yang dapat diketahui dari hasil evaluasi *confusion matrix*. Nilai persentase ini menunjukkan seberapa baik sebuah algoritma mampu mengklasifikasikan data. Hal ini menunjukkan seberapa berhasil algoritma mengklasifikasikan data pada akhirnya dalam proses klasifikasi. *Precision* adalah rasio yang tepat dari hasil prediksi positif model dibandingkan dengan total hasil prediksi positif. *Recall (sensitivity)*, mengukur kemampuan model untuk menemukan kasus yang benar-benar positif dalam keseluruhan data positif yang sebenarnya. Dengan mempertimbangkan total data positif yang ada, *recall* membantu dalam mengevaluasi sejauh mana model dapat menemukan dan memasukkan dengan benar kasus-kasus tersebut ke dalam hasil prediksi positifnya.

Proses validasi secara keseluruhan akan dimulai dengan menentukan data pelatihan dan pengujian sebelum diperoleh informasi tentang nilai akurasi, presisi dan *recall*. Cara membagi data secara keseluruhan

menjadi data pelatihan dan pengujian, *cross validation* akan menjadi tahap awal dalam evaluasi analisis klasifikasi. Dalam penelitian ini menggunakan *k-fold validation*.



Sumber: Penelitian (2023)

Gambar 5. *Cross Validation*

K-Fold cross-validation dalam klasifikasi jamur. Teknik *k-fold* ini bertujuan untuk mengatasi kelemahan metode *hold-out* dengan membagi dataset secara baru, mengatasi *"test only once bottleneck"*. Metode ini memungkinkan setiap observasi dari dataset asli memiliki kesempatan untuk muncul di kedua dataset, *training* dan *testing*. Ini menghasilkan model yang lebih tidak bias dan cocok untuk situasi di mana data masukan terbatas. Dengan penentuan ukuran *k-fold* yang telah ditentukan, k=8 digunakan sebagai langkah yang mendominasi dalam menentukan nilai akurasi paling tinggi.

Gambar 6 menunjukan nilai akurasi evaluasi yang dihasilkan. Berdasarkan tabel tersebut, nilai akurasi yang menunjukan hasil data pelatihan dan tes dengan pengukuran *confusion matrix* dalam proses klasifikasi jamur yang dapat dimakan dan jamur beracun, adalah 99.90%.

accuracy: 99.90% +/- 0.11% (micro average: 99.90%)

	true beracun	true dapat dimakan	class precision
pred. beracun	3908	0	100.00%
pred. dapat dimakan	8	4208	99.81%
class recall	99.80%	100.00%	

Sumber: Penelitian (2023)

Gambar 6. Nilai akurasi

Akurasi tersebut menunjukan hasil dari jumlah keseluruhan berdasarkan total jamur beracun (TP) dengan jumlah 3908 + 8 dan jamur yang dapat dimakan (TN) sebanyak 4208 dan dibagi dengan jumlah keseluruhan (TP + TN + FP + FN). Nilai akurasi 99,90% untuk klasifikasi data jamur berarti bahwa model klasifikasi berhasil memprediksi kelas jamur dengan benar sebanyak 99,90% dari total data yang digunakan serta menunjukkan bahwa model klasifikasi yang digunakan sangat akurat dan dapat diandalkan.

Setelah perhitungan melalui pengukuran persentase *confusion matrix* di dalam akurasi berdasarkan klasifikasi jamur, maka di dapat nilai presisi:

precision: 99.81% +/- 0.20% (micro average: 99.81%) (positive class: dapat dimakan)

	true beracun	true dapat dimakan	class precision
pred. beracun	3908	0	100.00%
pred. dapat dimakan	8	4208	99.81%
class recall	99.80%	100.00%	

Sumber: Penelitian (2023)

Gambar 7. Nilai presisi

Klasifikasi berhasil memprediksi kelas jamur yang dapat dikonsumsi dengan benar sebanyak 99,81% dari total jamur yang diprediksi sebagai jamur yang dapat dikonsumsi. Berikutnya disajikan juga hasil pengukuran *confusion matrix* berdasarkan tingkat presisi yang menghasilkan 99,81%. dalam Tabel 7 Nilai presisi.

Penentuan akhir yang disajikan setelah perhitungan ukuran tingkat presisi adalah *recall* atau *sensitivity*, Gambar 8. Menunjukan nilai *recall* yaitu hasil evaluasi identifikasi dalam hasil prediksi positif dimana 100.00% adalah presentase yang diperoleh.

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: dapat dimakan)

	true beracun	true dapat dimakan	class precision
pred. beracun	3908	0	100.00%
pred. dapat dimakan	8	4208	99.81%
class recall	99.80%	100.00%	

Sumber: Penelitian (2023)

Gambar 8. Nilai *recall*

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma C.45 dapat digunakan untuk mengklasifikasikan jamur dengan akurasi sebesar 99,90%. Berdasarkan interpretasi pohon keputusan, jamur yang memiliki : berbau amis, apak, busuk, creosof, menyengat, pedas, dan tidak berbau, warna cetakan spora hijau dan putih, akar tangkai klub dan bulat, ukuran insang sempit dan jarak insang dekat adalah jamur beracun. Sementara itu, jamur yang tidak memiliki karakteristik tersebut adalah jamur yang dapat dimakan. Penelitian ini dilakukan menggunakan software rapidminer pada dataset mushrooms dengan menggunakan algoritma C.45. Algoritma C.45 dapat menjadi metode yang efektif untuk membantu masyarakat dalam membedakan jamur yang dapat dimakan dan beracun. Dengan mengetahui karakteristik morfologi jamur yang dapat dimakan dan beracun, masyarakat dapat lebih berhati-hati dalam mengonsumsi jamur.

REFERENSI

Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. www.aaai.org

Haksoro, E. I., & Setiawan, A. (2022). Pengenalan Jamur yang Dapat Dikonsumsi Menggunakan Metode Transfer Learning pada Convolutional Neural Network. *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi Dan Komputer*, *5*(2), 81–91. https://doi.org/10.31961/eltikom.v5i2.428

Halling, R., Buchanan, P., Yun, W., & Cole, A. (2004). Edible and poisonous mushrooms of the world By Ian R. Hall, Steven L. Stephenson, Peter K. Buchanan, Wang Yun, Anthony L. J. Cole. 2003. *Brittonia*, *56*, 150. https://doi.org/10.1663/0007-196X(2004)056[0150:EAPMOT]2.0.CO;2

Hawksworth, D. L., & Rossman, A. Y. (1997). Where Are All the Undescribed Fungi? *Phytopathology®*, *87*(9), 888–891. https://doi.org/10.1094/PHYTO.1997.87.9.888

Ihsan, I., & Yusuf, F. (2013). *Analisis Jamur Beracun Berdasarkan Ciri Menggunakan Algoritma AdaBoost*. https://www.researchgate.net/publication/332440445_jurnal_KONIK_2013_Indah_Purwitasari_Ihsan?enrichId=rgreq-3799739def172eaa051431a6b51c16fa-XXX&enrichSource=Y292ZXJQYWdlOzMzMjQ0MDQ0NTtBUzo3NDgyMDA5MTA2ODAwNjZAMTU1NTM5NjQwNjEyOA%3D%3D&el=1_x_2&_esc=publicationCoverPdf

Khatun, S., Islam, A., Cakilcioglu, U., & Chatterjee, N. C. (2012). Research on Mushroom as a Potential Source of Nutraceuticals: A Review on Indian Perspective. In *Review Article American Journal of Experimental Agriculture* (Vol. 2, Issue 1). www.sciencedomain.org

Lancet. (1980). Mushroom Poisoning. *The Lancet*, *316*(8190), 351–352. https://doi.org/10.1016/S0140-6736(80)90346-3

Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *SESINDO 2015*, *2015*.

Nor Sarizan, F. H., Mustapha, M. F., Kairan, O., Mohd Bakhary, N. N., Azmira, N. H., & Ab Hamid, S. H. (2021). Analysis of identifying mushroom species using RapidMiner. *Journal of Mathematics and Computing Science (JMCS)*, *7*(2), 23–30.

Ortega, J., Lagman, A. C., Natividad, L. R. Q., Bantug, E. T., Resureccion, M. R., & Manalo, L. O. (2020). Analysis of performance of classification algorithms in mushroom poisonous detection using confusion matrix analysis. *International Journal*, *9*(1.3).

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R. B., & Campbell, N. A. (2014). *Campbell biology* (Tenth edition). Pearson Boston.

Rosela, Y. (2019). IMPLEMENTASI KLASIFIKASI DECISION TREE MENGANALISA STATUS PENJUALAN BARANG MENGGUNAKAN C4. 5 (Studi Kasus: Pt. Matahari Department Store Medan Mall). *Pelita Informatika: Informasi Dan Informatika*, *7*(3), 404–411.

Rudiyanto, A. R., Pujiono, P., Soeleman, M. A., & Mustagfirin, M. (2023). Performance of the Decision Tree Algorithm in the Classification of Edible and Poisonous Mushrooms with Information Gain Optimization. *Scientific Journal of Informatics*, *10*(4), 549–558.

Saputri, R. P. (n.d.). Penerapan Data Mining Technique sebagai Evaluasi Ketepatan Akurasi terhadap Klasifikasi Mushroom Data Set. *Institut Teknologi Sepuluh Nopember*.

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *11*(2), 612–619.

Ucha Putri, S., Irawan, E., Rizky, F., Tunas Bangsa, S., -Indonesia Jln Sudirman Blok No, P. A., & Utara, S. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5. In *Januari* (Vol. 2, Issue 1).

UCI Machine Learning Repository. (1981). *Mushroom*. UCI Machine Learning Repository. https://doi.org/https://doi.org/10.24432/C5959T

Wati, C. M., Fauzan, A. C., & Harliana, H. (2022). Performance Comparison Of Mushroom Type Classification Based On Multi-Scenario Dataset Using Decision Tree C4. 5 And C5. 0. *Jurnal Riset Informatika*, *4*(3), 247–258.

Zubair, A., & Muslikh, A. (2017). IDENTIFIKASI JAMUR MENGGUNAKAN METODE K-NEAREST NEIGHBOR DENGAN EKSTRAKSI CIRI MORFOLOGI. *Seminar Nasional Sistem Informasi (SENASIF)*, *1*(1). https://www.jurnalfti.unmer.ac.id/index.php/senasif/article/view/15