

Study On Identification Of Poisonous and Non-Toxic Mushrooms Using the Cart-Logitboost Algorithm

Studi tentang Identifikasi Jamur Beracun dan Tidak Beracun dengan Algoritma Cart-Logitboost

Moch. Anjas Aprihartha¹, Zulhandi Putrawan², Dicky Zulhan³, Fatma Ahardika Nurfaizal⁴

^{1,2,3,4} Program Studi PJJ Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Email: ¹anjas.aprihartha@dsn.dinus.ac.id, ²2118202200005@mhs.dinus.ac.id,

³3118202200011@mhs.dinus.ac.id, ⁴4118202300075@mhs.dinus.ac.id

Abstract

Mushrooms are one of the groups of living organisms in the fungal regnum which have umbrella-like body characteristics. The body consists of an upright part that functions as a rod to support the hood as well as a hood that is horizontal and rounded with different color variations. There are types of mushrooms that can be a food source for humans. Some types of mushrooms can be eaten or processed like other foods. Apart from that, some types of mushrooms are dangerous if consumed by humans because they are poisonous. Based on these problems, this study offers a new contribution in identifying types of poisonous and non-toxic mushrooms based on mushroom characteristics using the CART algorithm combined with the LogitBoost boosting algorithm. The aim of this research can be used as material for further studies in making tools that can effectively and accurately differentiate between poisonous and non-toxic types of mushrooms. This can help reduce cases of poisoning due to consumption of poisonous mushrooms. The data used is secondary data from public sources UCI Machine Learning Repository. Evaluation of model performance resulted in an accuracy of 98.79%; recall 98.70%; specificity 98.85%; precision 98.56%; F1-Score 98.63%, and AUC 0.9876. These results show that the model is very effective in detecting poisonous mushrooms and has minimal errors in classification.

Keywords: *CART, Mushrooms, LogitBoost, Poisonous, Non-Toxic*

Abstrak

Jamur merupakan salah satu kelompok jasad hidup dalam regnum fungi yang memiliki karakteristik tubuh seperti payung. Tubuhnya terdiri dari bagian tegak yang berfungsi sebagai batang menyangga tudung serta tudung yang berbentuk mendatar dan membulat dengan variasi warna berbeda. Terdapat jenis jamur yang dapat menjadi sumber makanan bagi manusia. Beberapa jenis jamur dapat dimakan atau diolah seperti bahan pangan lainnya. Selain itu, beberapa jenis jamur ada yang berbahaya jika dikonsumsi oleh manusia karena memiliki racun. Berdasarkan permasalahan tersebut, pada studi ini menawarkan kontribusi baru dalam melakukan identifikasi jenis jamur yang beracun dan tidak beracun berdasarkan karakteristik



JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika Nurfaizal

jamur dengan menggunakan algoritma CART yang dikombinasikan dengan algoritma penguat LogitBoost. Tujuan dari penelitian ini dapat dijadikan bahan kajian selanjutnya dalam pembuatan alat yang dapat secara efektif dan akurat membedakan jenis jamur beracun dan tidak beracun. Dengan demikian dapat membantu mengurangi kasus keracunan akibat konsumsi jamur beracun. Data yang digunakan adalah data sekunder dari sumber publik *UCI Machine Learning Repository*. Hasil penelitian diperoleh model terbaik dengan parameter optimal, yaitu rasio data adalah 90:10, kedalaman pohon (*maxdepth*) adalah 5, dan jumlah pohon (*ntree*) adalah 1000. Evaluasi kinerja model menghasilkan akurasi 98,79%; recall 98,70%; spesifisitas 98,85%; presisi 98,56%; F1-Score 98,63%, dan AUC 0,9876. Hasil ini menunjukkan model sangat efektif dalam mendeteksi jamur yang beracun dan minim kesalahan dalam klasifikasi.

Kata kunci: CART, Jamur, LogitBoost, Beracun, Tidak Beracun

1. PENDAHULUAN

Jamur merupakan salah satu kelompok jasad hidup dalam regnum fungi yang memiliki karakteristik tubuh seperti payung [14]. Tubuhnya terdiri dari bagian tegak yang berfungsi sebagai batang menyangga tudung serta tudung yang berbentuk mendatar dan membulat dengan variasi warna berbeda. Selain itu jamur memiliki keunikan yaitu memiliki sirip dibawah tudungnya dan dapat mengeluarkan spora atau darah seperti makhluk hidup lainnya.

Jamur dapat diselidiki pada dua aspek yang berlainan yaitu merusak dan berguna [1]. Terdapat jenis jamur yang dapat menjadi sumber makanan bagi manusia. Beberapa yang dapat dimakan atau diolah seperti bahan pangan lainnya. Jamur yang bisa dimakan tidak kurang dari 600 jenis, sekitar 35 jenis diantaranya dibudidayakan secara komersial. Namun, beberapa jenis jamur ada yang berbahaya jika dikonsumsi oleh manusia karena memiliki racun. Beberapa jenis jamur beracun memiliki rupa seperti jamur yang dapat dikonsumsi. Kemiripan ini yang biasa menjadi kesalahan orang-orang dalam mengidentifikasi sehingga jamur beracun disangka dapat dimakan. Akibatnya seseorang yang mengonsumsi jamur tersebut mengalami keracunan bahkan berisiko hingga kehilangan nyawa.

Berdasarkan permasalahan tersebut, pada studi ini akan dilakukan identifikasi jenis jamur yang beracun dan tidak beracun berdasarkan karakteristik jamur. Tujuan dari penelitian ini dapat dijadikan bahan kajian selanjutnya dalam pembuatan alat yang dapat secara efektif dan akurat membedakan jenis jamur beracun dan tidak beracun. Dengan demikian dapat membantu mengurangi kasus keracunan akibat konsumsi jamur beracun.

Beberapa penelitian sebelumnya membahas terkait klasifikasi jenis jamur beracun dan tidak beracun. Penelitian oleh Batubara, dkk. [6] yang mengklasifikasikan jamur beracun dan tidak beracun menggunakan algoritma Naive Bayes dan *K-Nearest Neighbour* (KNN). Penelitian oleh Wahdini [22] yang mengimplementasikan metode bayesian network dalam klasifikasi jamur beracun dan tidak beracun. Penelitian oleh Sahu, dkk. [18] membandingkan algoritma *Random Forest*, *Bagging*, *Gradient Boosting*, *Adaboost*, dan *Extreme Gradient Boosting* dalam klasifikasi jamur beracun dan tidak beracun. Kasus serupa dikaji oleh Tutuncu, dkk. [20] dengan membandingkan algoritma *Naive Bayes*, *Decision Tree*, *Support Vector Machine* dan *Adaboost*. Berdasarkan dua penelitian sebelumnya Sahu, dkk. [18] dan Tutuncu, dkk. [20] masing-masing memberikan kinerja model terbaik dengan algoritma *Adaboost*. *Adaboost* merupakan algoritma berulang yang melatih dan menggabungkan beberapa model lemah sehingga menjadi sebuah model yang

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

kuat sehingga dapat mencapai akurasi yang akurat [4][15]. Algoritma yang memiliki kemiripan dengan Adaboost dapat diidentifikasi sebagai algoritma LogitBoost karena keduanya memiliki karakteristik yang serupa yaitu berbasis *boosting*.

Studi kasus ini akan memberikan eksplorasi baru dalam identifikasi jenis jamur beracun dan tidak beracun dengan menggunakan algoritma klasifikasi seperti *Classification and Regression Tree* dengan LogitBoost (CART-LogitBoost). CART-LogitBoost adalah gabungan dari dua teknik berbeda. CART merupakan pengklasifikasi dasar yang menghasilkan model berupa pohon keputusan. Sedangkan LogitBoost adalah teknik peningkatan khusus yang dirancang untuk meningkatkan akurasi prediksi model dengan memberikan prioritas yang lebih besar pada contoh-contoh yang salah dikategorikan pada setiap iterasi, sehingga memperbaiki kesalahan dan memperkuat model. Algoritma LogitBoost digunakan untuk membangun model regresi logistik di setiap node pohon, kemudian digunakan pendekatan CART untuk pemangkasan [19]. Pemangkasan sangat penting untuk menghindari *overfitting*.

Penelitian oleh Cao, dkk. [7], menerapkan algoritma CART-LogitBoost dalam pengenalan beton keropos menggunakan pemerosesan gambar yang dioptimalkan secara metaheuristik. Hasil penelitian diperoleh tingkat akurasi 88,3%; presisi 88,9%; *recall* 87,4%; *score-F1* 88,1%; dan nilai prediksi negatif 87,4%. Penelitian oleh Alhalya, dkk. [2], perbandingan LogitBoost, Bagging, dan *support vector machine* (SVM) dalam segmentasi otomatis dan klasifikasi gambar thermal tangan di rheumatoid arthritis. Hasil penelitian diperoleh pengklasifikasi LogitBoost berkinerja lebih baik dengan akurasi 93,75% menggunakan teknik validasi silang 10 kali lipat dibandingkan dengan pengklasifikasi lainnya. Penelitian oleh Zandi, dkk [23] yang meneliti jaringan interaksi protein global dalam ragi *Saccharomyces cerevisiae* dan *Helicobacter pylori* dengan algoritma LogitBoost dan 10-*cross validation*. Hasil evaluasi dengan dataset mencapai akurasi 94,39% dan 97,89%. Penelitian lainnya oleh Fauzi, dkk. [11], membandingkan metode LogitBoost dan Bagging dalam mendeteksi penyakit kanker payudara dengan seleksi fitur berbasis *principal component analysis* (PCA). Berdasarkan tingkat akurasi, hasil prediksi dengan PCA berbasis LogitBoost sebesar 77.5862%, lebih baik dibandingkan dengan Bagging.

2. METODOLOGI PENELITIAN

2.1 Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan dataset yang berasal dari sumber publik *UCI Machine Learning Repository* [21]. Dataset terdiri dari 61069 sampel dengan berbagai variabel. Adapun variabel penelitian terdiri dari dua jenis yaitu variabel independen (X) dan variabel dependen (Y) yang disajikan pada Tabel 1.

Tabel 2.1. Variabel dan Jenis Data

Variabel	Jenis data
Kelas jamur (Y)	2 kategori
Diameter tudung jamur (X_1)	Numerik (cm)
Bentuk tudung jamur (X_2)	7 kategori
Permukaan tudung jamur (X_3)	10 kategori

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

Warna tudung jamur (X_4)	12 kategori
Apakah jamur memiliki memar atau mengeluarkan cairan? (X_5)	2 kategori
Cara lamela jamur menempel pada batang (X_6)	8 kategori
Jarak lamela jamur (X_7)	3 kategori
Warna lamela jamur (X_8)	13 kategori
Tinggi batang jamur (X_9)	Numerik (cm)
Lebar batang jamur (X_{10})	Numerik (cm)
Bentuk pangkal jamur (X_{11})	7 kategori
Permukaan batang jamur (X_{12})	11 kategori
Warna batang jamur (X_{13})	13 kategori
Jenis tirai jamur (X_{14})	2 kategori
Warna tirai jamur (X_{16})	13 kategori
Apakah jamur memiliki cincin? (X_{16})	2 kategori
Jenis cincin pada jamur (X_{17})	12 kategori
Warna Spora (X_{18})	12 kategori
Habitat (X_{19})	8 kategori
Musim tumbuh (X_{20})	4 kategori

2.2 Classification and Regression Tree (CART)

Classification and regression tree (CART) merupakan algoritma klasifikasi atau regresi yang menghasilkan model berupa pohon keputusan. Klasifikasi CART dilakukan dengan tiga tahapan yaitu membuat pohon klasifikasi dengan prosedur pembentukan menggunakan pemisahan simpul rekursif, memangkas pohon sehingga mendapatkan pohon yang lebih sederhana, dan menentukan pohon klasifikasi optimal [16]. CART dibangun dengan menggunakan kriteria Index Gini. Index Gini merupakan metode perhitungan yang ditujukan pada semua variabel independen dengan tujuan memberikan penilaian berdasarkan tingkat berpengaruhnya terhadap variabel dependen. Misalkan F merupakan himpunan data berisi n sampel dan m variabel independen dengan semua sampel i dinyatakan sebagai vektor $(x_1, x_2, x_3, \dots, x_m)$ dan salah satu kategori pada variabel dependen adalah c . Index Gini pada variabel j dapat dinyatakan sebagai berikut [3].

$$Gini(j) = 1 - \sum (P(c_k|x_j))^2 \quad (1)$$

dengan $P(c_k|x_j)$ merupakan peluang kelas c_k pada variabel independen ke x_j . Peluang $P(c_k|x_j)$ dapat dihitung melalui total sampel oleh kelas c_k dibagi dengan total keseluruhan sampel untuk setiap pemisahan.

2.3 LogitBoost

Friedman dkk. [12] memperkenalkan LogitBoost sebagai algoritma yang dibangun dengan meminimalkan Bernoulli log-likelihood melalui algoritma Newton adaptif untuk menyesuaikan model regresi logistik adaptif. Peningkatan Logit memperluas AdaBoost Diskrit dalam dua cara [9]. Pertama, ia menggunakan kemungkinan log Bernoulli sebagai

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

gantinya fungsi eksponensial sebagai fungsi kerugian. Selain itu, memperbarui pengklasifikasi dengan menambahkan model linier, bukan pembelajar biner yang lemah.

Diberikan $y^* = \frac{y+1}{2} \in \{0,1\}$ dengan $y = \{-1,1\}$. $F(x)$ merupakan fungsi pengklasifikasi gabungan dengan x tetap dan y merupakan variabel dependen kategori biner. Probabilitas prediksi dapat dicapai dengan

$$p(x) = \frac{e^{F(x)}}{e^{F(x)} + e^{-F(x)}} = \frac{1}{1 + e^{-2F(x)}} \quad (2)$$

Fungsi log-likelihood Bernaulli dinyatakan sebagai

$$l(y^*, p(x)) = -\ln(1 + e^{-2yF(x)}) \quad (3)$$

Menentukan z yang meminimalkan $l(y^*, p(x))$ dengan pendekatan Newton.

$$\begin{aligned} \frac{\partial l(y^*, p(x))}{\partial(F(x))} &= \frac{\partial(-\ln(1 + e^{-2yF(x)}))}{\partial(F(x))} \\ &= \frac{2ye^{-yF(x)}}{1 + e^{-2yF(x)}} \\ &= y^* - \frac{1}{1 + e^{-2F(x)}} \\ \frac{\partial^2 l(y^*, p(x))}{\partial^2(F(x))} &= \frac{\partial l}{\partial(F(x))} \left(\frac{2ye^{-yF(x)}}{1 + e^{-2yF(x)}} \right) \\ &= \left(1 - \frac{1}{1 + e^{-2yF(x)}} \right) \frac{1}{1 + e^{-2yF(x)}} \end{aligned} \quad (4)$$

Asumsikan probabilitas sukses ketika $y = 1$ maka persamaan (4) menjadi

$$\left(1 - \frac{1}{1 + e^{-2F(x)}} \right) \frac{1}{1 + e^{-2F(x)}} \quad (5)$$

Sehingga nilai z optimum diperoleh dengan

$$z = \frac{y^* - \frac{1}{1 + e^{-2F(x)}}}{\left(1 - \frac{1}{1 + e^{-2yF(x)}} \right) \frac{1}{1 + e^{-2yF(x)}}} \quad (6)$$

Jika $p(x) = \frac{1}{1 + e^{-2F(x)}}$ maka z dapat disederhanakan menjadi

$$z = \frac{y^* - p(x)}{(1 - p(x))p(x)} \quad (7)$$

dengan $(1 - p(x))p(x)$ menyatakan bobot w .

Diberikan pertimbangan pembaharuan $F(x) + f(x)$ dan ekspektasi log likelihood sebagai berikut.

$$E[l(F + f)] = E[2y^*(F(x) + f(x)) - \log[1 + e^{2(F(x)+f(x))}]] \quad (8)$$

Pada kondisi x bersyarat, akan ditentukan turunan pertama, $s(x)$ dan turunan kedua, $H(x)$ ketika $f(x) = 0$ pada persamaan (8).

$$\begin{aligned} s(x) &= \frac{\partial(E[l(F+f)])}{\partial f(x)} \Big|_{f(x)=0} \\ &= \frac{\partial(E[2y^*(F(x)+f(x)) - \log[1 + e^{2(F(x)+f(x))}])}{\partial f(x)} \Big|_{f(x)=0} \\ &= 2E \left(y^* - \frac{1}{1 + e^{-2F(x)}} \right) \\ s(x) &= 2E(y^* - p(x)|x), \\ H(x) &= \frac{\partial^2(E[l(F+f)])}{\partial^2 f(x)} \Big|_{f(x)=0} \\ &= \frac{\partial^2(E[2y^*(F(x)+f(x)) - \log[1 + e^{2(F(x)+f(x))}])}{\partial^2 f(x)} \Big|_{f(x)=0} \end{aligned} \quad (9)$$

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

$$= -4 \left[\frac{1}{1+e^{-2F(x)}} \left(1 - \frac{1}{1+e^{-2F(x)}} \right) \right]$$

$$H(x) = -4E(p(x)(1-p(x))|x) \quad (10)$$

dengan $p(x)$ didefinisikan dalam bentuk $F(x)$. Pembaruan Newton menjadi

$$F(x) \leftarrow F(x) - \frac{s(x)}{H(x)}$$

$$= F(x) - \frac{2E(y^*-p(x)|x)}{-4E(p(x)(1-p(x))|x)}$$

$$= F(x) + \frac{1}{2} E_w \left(\frac{y^*-p(x)}{p(x)(1-p(x))} \middle| x \right) \quad (11)$$

Secara ekuivalen, dengan pembaharuan newton $f(x)$ dapat menyelesaikan perkiraan kuadrat terkecil tertimbang ($F(x)$) ke fungsi log likelihood.

$$\min_{f(x)} E_{w(x)} \left(F(x) + \frac{1}{2} E_w \left(\frac{y^*-p(x)}{p(x)(1-p(x))} \right) - (F(x) - f(x)) \right) \quad (12)$$

2.4 Uji Performa Model

Confusion matrix merupakan alat yang efektif dalam menilai seberapa baik algoritma pengklasifikasi dalam mengidentifikasi pola dari berbagai kelas. *Confusion matrix* terdiri dari seperangkat baris dan kolom yang menampilkan hasil uji. model klasifikasi dalam bentuk tabel. Pada baris mencerminkan kelas-kelas aktual atau kelas observasi sedangkan pada kolom mewakili kelas-kelas prediksi. Akurasi adalah salah satu indikator yang paling banyak digunakan dalam klasifikasi kelas dan dalam perhitungannya mengambil data dari *confusion matrix* [8].

Tabel 2.2. *Confusion Matrix*

		Kelas Prediksi	
		Ya	Tidak
Kelas Observasi	Ya	a	b
	Tidak	c	d

Untuk mengetahui kinerja model dapat menggunakan alat ukur sebagai berikut [4][5][17].

$$Akurasi = \frac{a+d}{a+b+c+d} \times 100\% \quad (13)$$

$$Recall = \frac{a}{a+b} \times 100\% \quad (14)$$

$$Spesifisitas = \frac{d}{c+d} \times 100\% \quad (15)$$

$$Presisi = \frac{a}{a+c} \times 100\% \quad (16)$$

$$F1 \text{ Score} = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \times 100\% \quad (17)$$

$$AUC = \frac{1 + \left(\frac{a}{a+b} \right) - \left(1 - \frac{d}{c+d} \right)}{2} \quad (18)$$

Interpretasi kategori nilai AUC sebagai berikut [16].

- 0,90 – 1,00 = Klasifikasi Sempurna

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

- 0,80 – 0,89 = Klasifikasi Baik
- 0,70 – 0,79 = Klasifikasi Cukup Baik
- 0,60 – 0,69 = Klasifikasi Kurang Baik
- 0,50 – 0,59 = Klasifikasi Gagal

2.5 Tahapan Penelitian

Langkah-langkah dalam analisis data sebagai berikut:

- a. Membuang data yang tidak lengkap atau tidak konsisten sehingga dapat mengganggu proses analisis.
- b. Mendeskripsikan masing-masing variabel untuk melihat gambaran awal dataset yang dianalisis.
- c. Menganalisis data dengan melibatkan pengklasifikasi dasar yaitu CART dan proses *boosting* melibatkan teknik Logitboost dalam optimasi bobot optimal. Dalam analisis data memperhatikan parameter seperti rasio data *training* dan data *testing*, kedalaman data, dan jumlah pohon yang dibentuk.
- d. Menguji model akhir menggunakan data *testing*.
- e. Menginterpretasi hasil.

3. HASIL DAN PEMBAHASAN

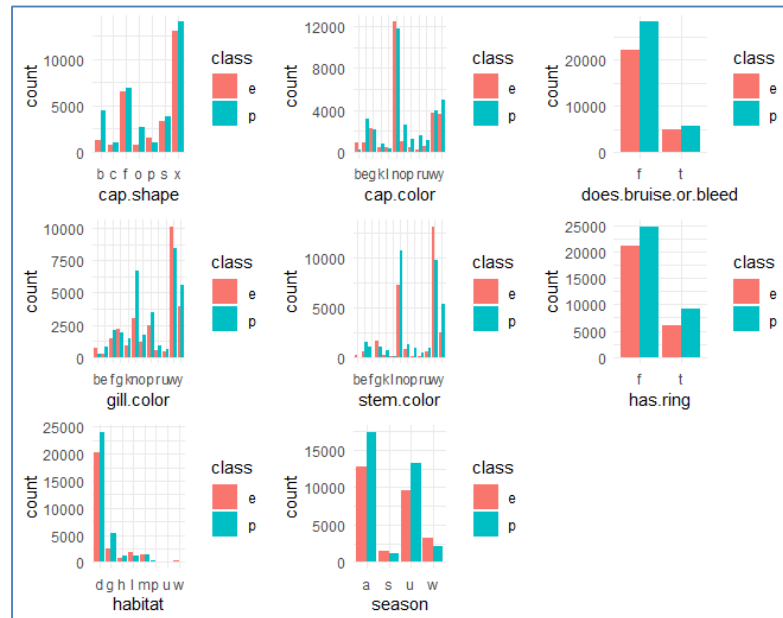
3.1 Eksplorasi Data

Hasil investigasi data yang diperoleh dari *UCI Machine Learning Repository* [21] ditemukan beberapa variabel yang mengandung data yang hilang (*missing value*). Variabel yang terdapat *missing value* yaitu permukaan tudung jamur, cara lamela jamur menempel pada batang, jarak lamela jamur, bentuk pangkal jamur, permukaan batang jamur, jenis tirai jamur, warna tirai jamur, jenis cincin pada jamur, warna spora. Oleh karena itu, variabel tersebut diabaikan dalam penelitian ini karna dapat mengganggu proses analisis selanjutnya.

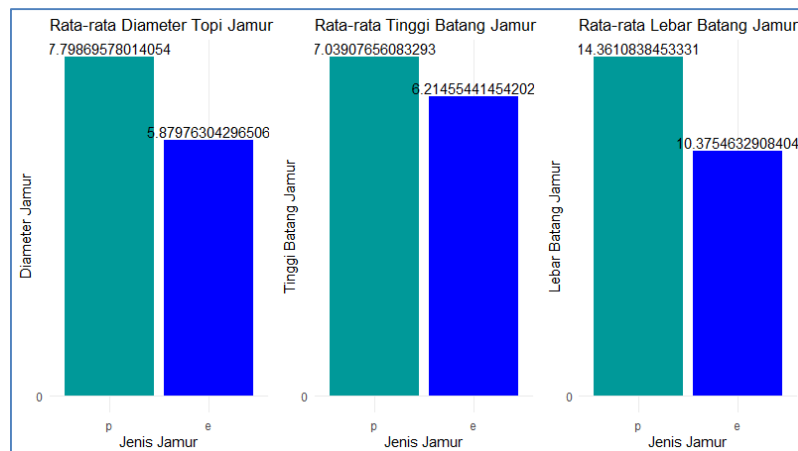
Pada Gambar 3.1. Grafik bentuk tudung jamur menunjukkan bahwa bentuk tudung *convex* (*x*) paling banyak ditemukan 13974 jamur beracun, sedangkan bentuk tudung *conical* (*c*) paling sedikit ditemukan 774 jamur tidak beracun. Grafik warna tudung jamur menunjukkan bahwa warna tudung coklat paling banyak ditemukan pada jamur beracun dan tidak beracun dengan masing-masing 11811 amatan dan 12407 amatan. Grafik jamur memar atau mengeluarkan cairan (*does bruise or bleed*) menunjukkan bahwa ada 28240 jamur beracun yang memar atau mengeluarkan cairan sedangkan ada 4942 jamur tidak beracun yang tidak memar atau mengeluarkan cairan. Grafik warna lamela jamur (*gill color*) menunjukkan bahwa warna lamela putih (*w*) paling banyak ditemukan pada jamur tidak beracun dibandingkan warna lamela merah (*e*). Hal yang sama ditemukan pada grafik warna batang jamur (*stem color*), batang berwarna putih paling banyak ditemukan pada jamur tidak beracun sedangkan batang berwarna coklat tidak ditemukan jamur tidak beracun, demikian pula batang berwarna biru tidak ditemukan jamur yang beracun. Grafik jamur yang memiliki cincin (*has ring*) menunjukkan bahwa jamur yang memiliki cincin lebih banyak pada jamur beracun dibandingkan yang tidak beracun. Grafik habitat jamur menunjukkan ada 24046 jamur beracun ditemukan di pohon (*d*) sedangkan tidak ditemukan

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

jamur beracun di limbah (w) atau perkotaan (u). Grafik musim ($season$) menunjukkan bahwa ditemukan jamur beracun dan tidak beracun paling banyak pada musim gugur (a). Pada musim gugur ditemukan 17392 jamur beracun dan 12785 jamur tidak beracun.



Gambar 3.1. Asosiasi Variabel Independen Kategorik dan Variabel Dependen



Gambar 3.2. Asosiasi Variabel Independen Numerik dan Variabel Dependen

Pada Gambar 3.2 diperlihatkan bahwa diameter tudung, tinggi dan lebar batang jamur beracun lebih besar dibandingkan dengan jamur tidak beracun. Grafik diameter jamur menunjukkan bahwa jamur beracun memiliki rata-rata diameter tudung 7,90 cm lebih besar dibandingkan jamur tidak beracun yaitu 5,88 cm. Grafik tinggi batang jamur menunjukkan bahwa jamur beracun memiliki rata-rata tinggi batang 7,04 cm lebih tinggi dibandingkan jamur tidak beracun yaitu 6,21 cm. Grafik lebar batang jamur menunjukkan bahwa jamur

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

beracun memiliki rata-rata lebar batang 14,36 cm lebih besar dibandingkan jamur tidak beracun yaitu 10,36 cm.

3.2 Evaluasi Model CART-LogitBoost

Tahap awal akan dilakukan pemodelan CART-Logitboost dengan parameter yang telah ditentukan secara *default*, yaitu pada kedalaman maksimum setiap pohon sama dengan tiga ($maxdepth = 3$) dan jumlah pohon adalah 100 ($ntree = 100$), untuk memilih rasio data *training* dan data *testing* yang terbaik berdasarkan *out of bag* (OOB) *error rate*. estimasi OOB dilakukan dengan mengambil amatan secara bootstrap sehingga terbentuk sampel *bootstrap*, beberapa amatan yang tidak terambil disebut *out of bag data* [10]. Pada sampel *out of bag* data dibangun pohon-pohon keputusan yang selanjutnya digunakan untuk prediksi sampel data *training*. *Out of bag error rate* diperoleh dengan menghitung rata-rata tingkat kesalahan prediksi pada setiap pohon dari sampel *out of bag* data. Hasil uji disajikan pada Tabel 3.1.

Tabel 3.1. OOB *Error Rate* Terhadap Proporsi Data

Proporsi Data	OOB
60:40	0,174
70:30	0,174
80:20	0,182
90:10	0,173

Pada Tabel 3.1 memperlihatkan nilai OOB *error rate* cukup rendah, namun proporsi data 90:10 jauh lebih baik dengan OOB *error rate* paling rendah sebesar 0,173. Ini menunjukkan data *training* 90% dan data *testing* 10% dapat meningkatkan performa model CART-LogitBoost.

Tahap kedua memilih kedalaman maksimum setiap pohon ($maxdepth$), dengan rasio data *training* dan data *testing*, yaitu 90:10 dan parameter $ntree$ yang ditentukan secara *default* yaitu 100 pohon. Hasil uji model ditampilkan pada Tabel 3.2.

Tabel 3.2. OOB *Error Rate* Terhadap Kedalaman Pohon

<i>Maxdepth</i>	OOB
1	0,336
2	0,241
3	0,173
4	0,160
5	0,141
6	0,164
7	0,176
8	0,182
9	0,182
10	0,180

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

Berdasarkan Tabel 3.2, Nilai OOB *error rate* mengalami penurunan dari *maxdepth* 1 sampai 5 kemudian terjadi peningkatan ketika *maxdepth* 6. Kedalaman level optimal pada *maxdepth* 5 yang memberikan OOB *error rate* terendah sebesar 0,141, sehingga kedalaman level ini adalah paling optimal dalam mengurangi terjadinya *overfitting*.

Tahap ketiga memilih jumlah pohon (*n_{tree}*), dengan rasio data *training* dan data *testing*, yaitu 90:10 dan *maxdepth* 5. Hasil uji disajikan pada Tabel 3.3.

Tabel 3.3. OOB *Error Rate* Terhadap Jumlah Pohon

<i>Ntree</i>	OOB
100	0,141
200	0,132
300	0,087
400	0,067
500	0,058
600	0,037
700	0,037
800	0,032
900	0,028
1000	0,023

Pada Tabel 3.3 diperlihatkan bahwa semakin meningkatnya jumlah pohon maka OOB *error rate* menurun secara konsisten. Hal ini berarti bertambahnya pohon akan meningkatkan performa model. Selain itu dengan lebih banyak pohon akan mengurangi varian dan dapat meningkatkan akurasi prediksi. Nilai OOB *error rate* terus menurun hingga jumlah pohon mencapai 1000 dengan menghasilkan OOB *error rate* sebesar 0,023.

Pada tiga tahap yang telah dilakukan maka model terbaik diperoleh dengan parameter optimal, yaitu rasio data adalah 90:10, kedalaman pohon (*maxdepth*) adalah 5, dan jumlah pohon (*n_{tree}*) adalah 1000. Dengan parameter-parameter ini, model akan mencapai keseimbangan terbaik dalam hal kompleksitas dan kemampuan generalisasi dalam memprediksi data baru.

Selanjutnya dilakukan pengujian model dengan data *testing* untuk menunjukkan hasil kemampuan model pada parameter yang terpilih. Hasil uji disajikan dalam bentuk *confusion matrix* pada Tabel 3.4.

Tabel 3.4. *Confusion Matrix* Model CART-LogitBoost

		Kelas Prediksi	
		Beracun	Tidak Beracun
Kelas Aktual	Beracun	2664	35
	Tidak Beracun	39	3369

Pada Tabel 3.4, Jumlah kasus yang terprediksi tepat sebagai jamur beracun sebanyak 2664 amatan dan jumlah kasus yang terprediksi tepat sebagai jamur tidak beracun sebanyak

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

3369 amatan. Sisanya 74 amatan adalah missklasifikasi. Hasil evaluasi kinerja model CART-LogitBoost ditunjukkan pada Tabel 3.5.

Tabel 3.5. Evaluasi Kinerja Model LogitBoost Pada Data *Testing*

Pengukuran	Hasil
Akurasi	98,79%
Recall	98,70%
Spesifisitas	98,85%
Presisi	98,56%
F1-Score	98,63%
AUC	0,9876

Akurasi yang dihasilkan sebesar 98,79% yang berarti model berhasil mengklasifikasikan jamur beracun dan tidak beracun dengan tepat dalam 98,79% kasus. Recall sebesar 98,70% yang berarti model berhasil mengidentifikasi jamur yang beracun dengan tepat sebesar 98,70%. Spesifisitas 98,85% yang berarti model berhasil mengidentifikasi jamur yang tidak beracun dengan tepat sebesar 98,85%. Presisi sebesar 98,56% yang berarti semua jamur yang diklasifikasikan sebagai beracun, sebesar 98,56% memang benar-benar beracun. F1-Score sebesar 98,63% menunjukkan kemampuan model yang sangat baik dalam mendeteksi jamur yang beracun. Nilai AUC sebesar 0,9876 menunjukkan kemampuan model hampir sempurna dalam mengklasifikasikan jamur yang beracun dan jamur yang tidak beracun. Tingginya persentase akurasi, recall, spesifisitas, presisi, F1-Score, dan AUC menunjukkan model sangat efektif dalam mendeteksi jamur yang beracun dan minim kesalahan dalam klasifikasi.

Penelitian dilakukan oleh Batubara, dkk. [6] yang mengklasifikasikan jamur beracun dan tidak beracun menggunakan algoritma Naive Bayes dan K-Nearest Neighbour (KNN). Pada algoritma Naive Bayes, akurasi, recall, dan presisi masing-masing sebesar 92%, 91,5%, dan 92,5%. Pada algoritma KNN, akurasi, recall, dan presisi masing-masing sebesar 98%, 98%, dan 92,5%. Penelitian lainnya oleh Wahdini [22] yang mengimplementasikan metode bayesian network dalam klasifikasi jamur beracun dan tidak beracun. Hasil evaluasi model dengan data *testing* menghasilkan akurasi sebesar 98%. Berdasarkan penelitian sebelumnya maka algoritma CART-LogitBoost memberikan kemampuan klasifikasi yang lebih baik dengan akurasi yang dihasilkan sebesar 98,79%.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan, dari dataset yang berjumlah 61069 sampel dengan sebelas variabel independen dan sebuah variabel dependen kategori biner akan dilakukan pemodelan dengan algoritma *decision tree* tipe CART yang dikombinasikan dengan algoritma LogitBoost. Model CART-LogitBoost dibangun dengan tiga tahapan pengujian parameter untuk menghasilkan model terbaik. Setelah tiga tahap pengujian dilakukan diperoleh model terbaik dengan parameter optimal, yaitu rasio data adalah 90:10, kedalaman pohon (*maxdepth*) adalah 5, dan jumlah pohon (*ntree*) adalah 1000. Kemudian pada model tersebut dilakukan pengujian dengan data *testing*, dan menghasilkan persentase

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

akurasi, recall, spesifisitas, presisi, F1-Score yang sangat baik. Uji AUC diperoleh sebesar 0,9876 yang menunjukkan model sangat efektif dalam mendeteksi jamur yang beracun dan minim kesalahan dalam klasifikasi.

DAFTAR PUSTAKA

- [1] Achmad, I., Mugiono, S. P., Tias Arlianti, S. P., & Chotimatul Azmi, S. P., 2011. *Panduan Lengkap Jamur*. Penebar Swadaya Grup.
- [2] Ahalya, R. K., Snehalatha, U., & Dhanraj, V., 2023. Automated Segmentation and Classification of Hand Thermal Images in Rheumatoid Arthritis Using Machine Learning Algorithms: A Comparison with Quantum Machine Learning Technique. *Journal of Thermal Biology*, Vol. 111, 103404. <https://doi.org/10.1016/j.jtherbio.2022.103404>
- [3] Aprihartha, M. A., Astutik, F., & Sulistianingsih, N., 2024. Comparison of Naïve Bayes, CART, dan CART Adaboost Methods in Predicting Tire Product Sales. *Jurnal Matematika, Statistika dan Komputasi*, Vol. 20, No. 3, 596-605. <https://doi.org/10.20956/j.v20i3.33187>
- [4] Aprihartha, M. A. Implementasi CART-Real Adaboost dalam Memprediksi Minat Pelanggan Membeli Sepatu. *Jurnal EurekaMatika*, 12(1), 35-46. <https://doi.org/10.17509/jem.v12i1.67808>
- [5] Aprihartha, M. A., Alam, T. N., & Husniyadi, M. (2024). Perbandingan Metrik Euclidean dan Metrik Manhattan untuk K-Nearest Neighbors dalam Klasifikasi Kismis. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 21-30. <https://doi.org/10.54082/jiki.126>
- [6] Batubara, G. M. C., Desiani, A., & Amran, A., 2023. Klasifikasi Jamur Beracun Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors. *Jurnal Ilmu Komputer dan Informatika*, Vol. 3, No. 1, 33-42. <https://doi.org/10.54082/jiki.68>
- [7] Cao, M. T., Nguyen, N. M., Chang, K. T., Tran, X. L., & Hoang, N. D., 2021. Automatic Recognition of Concrete Spall Using Image Processing and Metaheuristic Optimized LogitBoost Classification Tree. *Advances in Engineering Software*, Vol. 159, 103031. <https://doi.org/10.1016/j.advengsoft.2021.103031>
- [8] Chacón, A. M. P., Ramírez, I. S., & Márquez, F. P. G., 2023. K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection. *Sustainable Futures*, Vol. 6, 100132. <https://doi.org/10.1016/j.advengsoft.2021.103031>
- [9] Chu, J., Lee, T. H., & Ullah, H., 2018. Component-wise AdaBoost Algorithms for High-dimensional Binary Classification and Class Probability Prediction. Department of Economics, University of California. <https://doi.org/10.1016/bs.host.2018.10.003>
- [10] Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157-175. https://doi.org/10.1007/978-1-4419-9326-7_5
- [11] Fauzi, A., Supriyadi, R., & Maulidah, N., 2020. Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest. *Jurnal Infortech*, Vol. 2, No. 1, 96-101. <https://doi.org/10.31294/infortech.v2i1.8079>
- [12] Friedman, J., Hastie, T., & Tibshirani, R., 2000. Additive Logistic Regression: A Statistical View of Boosting. *Annals of Statistics*, Vol. 28, No. 2, 337-374. DOI: 10.1214/aos/1016218223
- [13] Gorunescu, F., 2011. *Data Mining: Concepts, Models and Techniques*. Springer Science & Business Media, Berlin.
- [14] Hasanuddin, H., 2018. Jenis Jamur Kayu Makroskopis Sebagai Media Pembelajaran Biologi (Studi di TNGL Blangjerango Kabupaten Gayo Lues). BIOTIK: *Jurnal Ilmiah Biologi Teknologi dan Kependidikan*, Vol. 2, No. 1, 38-52. <http://dx.doi.org/10.22373/biotik.v2i1.234>

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Moch. Anjas Aprihartha, Zulhandi Putrawan, Dicky Zulhan, Fatma Ahardika
Nurfaizal

- [15] Hu, G., Yin, C., Wan, M., Zhang, Y., & Fang, Y., 2020. Recognition of diseased pinus trees in UAV images using deep learning and AdaBoost classifier. *Biosystems Engineering*, 194, 138-151.
- [16] Kuswanto, H., & Mubarak, R., 2019. Classification of Cancer Drug Compounds for Radiation Protection Optimization Using CART. *Procedia Computer Science*, 161, 458-465. <https://doi.org/10.1016/j.procs.2019.11.145>
- [17] Prasetya, J., Fallo, S. I., & Aprihartha, M. A., 2024. Stacking Machine Learning Model for Predict Hotel Booking Cancellations. *Jurnal Matematika, Statistika dan Komputasi*, 20(3), 525-537. <https://doi.org/10.20956/j.v20i3.32619>
- [18] Sahu, R., Pandey, S., Verma, R., & Pandey, P., 2024. Ensemble Learning based Classification of Edible and Poisonous Agaricus Mushrooms. In *2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-7). IEEE. DOI: [10.1109/ICAECT60202.2024.10469539](https://doi.org/10.1109/ICAECT60202.2024.10469539)
- [19] Sarkar, S. K., Rudra, R. R., Talukdar, S., Das, P. C., Nur, M. S., Alam, E., ... & Islam, A. R. M. T., 2024. Future Groundwater Potential Mapping Using Machine Learning Algorithms and Climate Change Scenarios in Bangladesh. *Scientific Reports*, Vol. 14, No. 1, 10328. <https://doi.org/10.1038/s41598-024-60560-2>
- [20] Tutuncu, K., Cinar, I., Kursun, R., & Koklu, M., 2022. Edible and poisonous mushrooms classification by machine learning algorithms. In *2022 11th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-4). IEEE. DOI: [10.1109/MECO55406.2022.9797212](https://doi.org/10.1109/MECO55406.2022.9797212)
- [21] Wagner, Dennis, Heider, D., and Hattab, Georges, 2023. *Secondary Mushroom*. UCI Machine Learning Repository. <https://doi.org/10.24432/C5FP5Q>.
- [22] Wahdini, M. G., 2020. Implementasi Metode Bayesian Network untuk Sistem Rekomendasi Klasifikasi Jamur Beracun dan Tidak Beracun. Disertasi Doktorat, Universitas Hasanuddin.
- [23] Zandi, F., Mansouri, P., & Goodarzi, M., 2023. Global Protein-Protein Interaction Networks in Yeast *Saccharomyces Cerevisiae* and *Helicobacter Pylori*. *Talanta*, 265, 124836. <https://doi.org/10.1016/j.talanta.2023.124836>