

LAPORAN TUGAS BESAR

KECERDASAN BUATAN

Mushroom Classification using Decision Tree

Robby Hidayah Ramadhan ¹⁾, Khoirunnisa ²⁾, Fadia Dilla Sabine ³⁾

Muhammad Rasyid Aditya ⁴⁾, Heggie Ilham ⁵⁾

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

Email : robby.120450033@student.itera.ac.id ¹⁾, khoirunnisa.120450029@student.itera.ac.id ²⁾,

fadia.120450011@student.itera.ac.id ³⁾, muhammad.120450089@student.itera.ac.id

⁴⁾, heggie.120450005@student.itera.ac.id ⁵⁾

Abstrak

Jamur atau fungi saat ini banyak berperan penting dalam kehidupan manusia, jamur bisa dijadikan sebagai tanaman hias, produk industri, serta dapat dikonsumsi oleh manusia. Banyak spesies jamur yang hampir serupa dan beberapa diantaranya adalah jamur beracun. Untuk alasan itu, orang harus berhati-hati untuk mengkonsumsi beberapa jamur karena tidak semua jamur dapat dikonsumsi oleh manusia. Identifikasi jamur yang salah bisa menjadi masalah besar karena jika seseorang terus mengkonsumsi jamur tanpa mengetahui jenis jamur yang aman atau tidak. Maka dibuat pemodelan untuk klasifikasi jenis-jenis jamur menggunakan Decision Tree dengan jenis algoritma CART untuk mempermudah dalam membedakan jenis-jenis jamur yang ditemui. Data yang digunakan pada penelitian ini merupakan *Mushroom Classification* yang merupakan data sekunder. Dari hasil analisis, diperoleh jamur edible (jamur yang dapat dikonsumsi) memiliki jumlah sebanyak 4,208 dan poisonous (jamur beracun) memiliki jumlah 3,916 dari keseluruhan data. Dari pemodelan yang dibuat, diperoleh nilai akurasi tertinggi yaitu 99,7%.

Kata Kunci: Decision Tree, *Edible*, Jamur, *Poisonous*

Abstract

Mushrooms or fungi currently play many important roles in human life, mushrooms can be used as ornamental plants, industrial products, and can be consumed by humans. Many species of mushrooms are almost similar and some of them are poisonous mushrooms. For that reason, one has to be careful about consuming some mushrooms as not all mushrooms are edible by humans. Identification of the wrong mushroom can be a big problem because if someone continues to consume mushrooms without knowing which type of mushroom is safe or not. Then a model was created for the classification of mushroom types using a Decision Tree with the CART algorithm type to make it easier to distinguish the types of mushrooms encountered. The data used in this study is Mushroom Classification which is secondary data. From the results of the analysis, it was found that edible mushrooms (mushrooms that can be consumed) have a total of 4,208 and poisonous mushrooms have a total of 3,916 from the total data. From the modeling that was made, the highest accuracy value was obtained, namely 99.7%.

Keywords: Decision Tree, *Edible*, Identification, Mushrooms, *Poisonous*

I. PENDAHULUAN

Saat ini ada berbagai macam jenis jamur di dunia. Jamur atau fungi saat ini banyak berperan penting dalam kehidupan manusia, jamur bisa dijadikan sebagai tanaman hias, produk industri, serta dapat dikonsumsi oleh manusia. Sebagai makanan manusia, artinya ada beberapa jenis jamur yang layak dan aman untuk dikonsumsi oleh manusia. Jamur tidak hanya memiliki 1 jenis, jamur memiliki banyak *genus* dan *famili* yang berbeda-beda. Banyak spesies jamur yang hampir serupa dan beberapa diantaranya adalah jamur beracun. Untuk alasan itu, orang harus berhati-hati untuk mengonsumsi beberapa jamur karena tidak semua jamur dapat dikonsumsi oleh manusia.

Diperkirakan sebanyak 140.000 jenis jamur telah diinventarisasi dari seluruh dunia dimana kurang lebih 2000 jenis merupakan jamur yang dapat dikonsumsi, namun sebagian jenis lainnya merupakan kelompok beracun (tidak diketahui dengan jelas jumlahnya) karena menghasilkan toksin [1]. Sebagian besar kasus keracunan jamur disebabkan oleh kesalahan identifikasi atau pengenalan jamur yang dikonsumsi. Hal tersebut terjadi karena karakter morfologi jamur yang kompleks dan seringkali terlihat mirip antara satu dengan lainnya.

Berdasarkan beberapa liputan berita, ada beberapa kasus terjadi dimana orang mengalami keracunan dan meninggal dunia karena mengonsumsi jamur beracun. Sebagai contoh pertama, terdapat kasus di Afghanistan dimana salah satu anak dari keluarga di Afghanistan yang harus dievaluasi setelah makan sup yang mengandung *cap mushroom* atau jamur topi yang mereka dapat dari mengumpulkan jamur-jamur di hutan Polandia di luar pusat karantina mereka. Kasus lain adalah seorang wanita yang merupakan pekerja ASHA¹ meninggal sementara 3 anggota keluarga lainnya dirawat di rumah sakit terdekat setelah menelan jamur. Sedangkan menurut data yang telah dilakukan terdapat 76 kasus keracunan jamur liar dalam kurun waktu 10 tahun (2010-2020) di berbagai wilayah di Indonesia. Dari 550 orang yang menjadi korban 9 diantaranya tewas karena mengonsumsi jamur beracun. Data menunjukkan bahwa setiap tahunnya selalu terjadi kasus keracunan jamur liar [2].

¹ American School Health Association

II. TUJUAN

Identifikasi jamur yang salah bisa menjadi masalah besar karena jika seseorang terus mengkonsumsi jamur tanpa mengetahui jenis jamur yang aman atau tidak karena dapat menyebabkan kematian yang mana dapat meningkatkan kasus kematian karena jamur beracun. Namun, banyaknya jenis jamur membuat masyarakat awam sulit untuk membedakan jenis-jenis jamur. Maka dibuat pemodelan untuk klasifikasi jenis-jenis jamur menggunakan *Decision Tree* dengan jenis algoritma CART untuk mempermudah dalam membedakan jenis-jenis jamur yang ditemui.

III. METODE

Percobaan yang dilakukan menggunakan Decision Tree dengan jenis algoritma CART (*Classification and Regression Tree*). Decision Tree merupakan pendekatan pemodelan prediktif. Decision tree digunakan untuk mengatasi permasalahan dalam klasifikasi, *data mining*, serta *machine learning*. Sementara, algoritma CART merupakan salah satu jenis algoritma Decision Tree. CART adalah algoritma prediktif yang digunakan untuk menjelaskan bagaimana variabel target dapat diprediksi. CART mudah dipahami dan divisualisasikan.

Adapun kelebihan dan kekurangan Decision Tree adalah :

Kelebihan	Kekurangan
<ul style="list-style-type: none">- Tidak perlu preprocessing data- Tidak ada asumsi tentang distribusi data- Mengelola kolinearitas secara efisien- Pohon keputusan dapat memberikan penjelasan yang jelas untuk ramalan tersebut	<ul style="list-style-type: none">- Jika kita memaksa untuk membangun model decision tree yang high purity maka model akan kecenderungan dapat mengalami overfitting- Model rawan outlier- Semakin kompleks dataset maka decision tree yang dibangun pun akan semakin kompleks

3.1 DATA

Dataset yang digunakan mencakup deskripsi sampel hipotesis yang sesuai dengan 23 spesies jamur. Setiap spesies diidentifikasi sebagai jamur yang bisa dimakan, jamur yang beracun, dan

jamur yang tidak dapat dimakan atau direkomendasikan. Kelas yang terakhir (jamur yang tidak bisa dimakan dan tidak direkomendasikan) digabungkan ke dalam kelas yang beracun. Terdapat beberapa kolom dengan informasi sebagai berikut.

- 1) Class, menyatakan kelas pada dataset.
Keterangan: *edible* = *e*, *poisonous* = *p*
- 2) Cap-shape, menyatakan bentuk bagian atas jamur (topi jamur).
Keterangan: *bell* = *b*, *conical* = *c*, *convex* = *x*, *flat* = *f*, *knobbed* = *k*, *sunken* = *s*
- 3) Cap-surface, menyatakan bentuk permukaan topi jamur.
Keterangan: *fibrous* = *f*, *grooves* = *g*, *scaly* = *y*, *smooth* = *s*
- 4) Cap-color, menyatakan warna dari topi jamur.
Keterangan: *brown* = *n*, *buff* = *b*, *cinnamon* = *c*, *gray* = *g*, *green* = *r*, *pink* = *p*, *purple* = *u*, *red* = *e*, *white* = *w*, *yellow* = *y*
- 5) Bruises, menyatakan corak/memar tertentu pada jamur.
Keterangan: *bruises* = *t*, *no* = *f*
- 6) Odor, menyatakan bau dari jamur.
Keterangan: *almond* = *a*, *anise* = *l*, *creosote* = *c*, *fishy* = *y*, *foul* = *f*, *musty* = *m*, *none* = *n*, *pungent* = *p*, *spicy* = *s*.
- 7) Gill-attachment, gill merupakan bagian yang berada di bawah topi jamur. Gill-attachment menyatakan jenis atau bentuk dari bagian bawah topi jamur (gill).
Keterangan: *attached* = *a*, *descending* = *d*, *free* = *f*, *notched* = *n*.
- 8) Gill-spacing, menyatakan jarak pada bagian bawah topi jamur (gill).
Keterangan: *close* = *c*, *crowded* = *w*, *distant* = *d*.
- 9) Gill-size, menyatakan ukuran pada bagian bawah topi jamur (gill).
Keterangan : *broad* = *b*, *narrow* = *n*
- 10) Gill-color, menyatakan warna pada bagian bawah topi jamur (gill).
Keterangan: *black* = *k*, *brown* = *n*, *buff* = *b*, *chocolate* = *h*, *gray* = *g*, *green* = *r*, *orange* = *o*, *pink* = *p*, *purple* = *u*, *red* = *e*, *white* = *w*, *yellow* = *y*
- 11) Stalk-shape
Keterengnan: *enlarging* = *e*, *tapering* = *t*
- 12) Stalk-root
Keterangan: *bulbous* = *b*, *club* = *c*, *cup* = *u*, *equal* = *e*, *rhizomorphs* = *z*, *rooted* = *r*

13) Stalk-surface-above-ring

Keterangan: *fibrous* = *f*, *scaly* = *y*, *silky* = *k*, *smooth* = *s*

14) Stalk-surface-below-ring

Keterangan: *fibrous* = *f*, *scaly* = *y*, *silky* = *k*, *smooth* = *s*

15) Stalk-color-above-ring

Keterangan: *brown* = *n*, *buff* = *b*, *cinnamon* = *c*, *gray* = *g*, *orange* = *o*, *pink* = *p*, *red* = *e*,
white = *w*, *yellow* = *y*

16) Stalk-color-below-ring

Keterangan: *brown* = *n*, *buff* = *b*, *cinnamon* = *c*, *gray* = *g*, *orange* = *o*, *pink* = *p*, *red* = *e*,
white = *w*, *yellow* = *y*

17) Veil-type

Keterangan: *partial* = *p*, *universal* = *u*

18) Veil-color

Keterangan: *brown* = *n*, *orange* = *o*, *white* = *w*, *yellow* = *y*

19) Ring-number

Keterangan: *none* = *n*, *one* = *o*, *two* = *t*

20) Ring-type

Keterangan: *cobwebby* = *c*, *evanescent* = *e*, *flaring* = *f*, *large* = *l*, *none* = *n*, *pendant* = *p*,
sheathing = *s*, *zone* = *z*

21) Spore-print-color

Keterangan: *black* = *k*, *brown* = *n*, *buff* = *b*, *chocolate* = *h*, *green* = *r*, *orange* = *o*, *purple*
= *u*, *white* = *w*, *yellow* = *y*

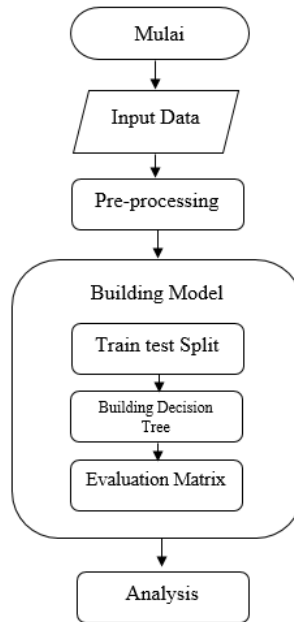
22) Population

Keterangan: *abundant* = *a*, *clustered* = *c*, *numerous* = *n*, *scattered* = *s*, *several* = *v*, *solitary*
= *y*

23) Habitat

Keterangan: *grasses* = *g*, *leaves* = *l*, *meadows* = *m*, *paths* = *p*, *urban* = *u*, *waste* = *w*, *woods*
= *d*

3.2 Tahapan Metode



Gambar 1. Flowchart System

Terdapat beberapa tahapan yang diterapkan sebagai bagian dari metode pembelajaran yang dilakukan seperti yang terdapat pada Gambar 1.

3.3 Input Data

Data yang digunakan pada penelitian ini merupakan *Mushroom Classification*. Dataset yang digunakan merupakan data publik atau data sekunder. Dataset ini bukan berasal dari penelitian langsung, melainkan dari sumber lain. Dataset *Mushroom Classification* diambil dari situs Kaggle, di mana Kaggle merupakan situs yang dibuat oleh Anthony Goldbloom untuk berlomba membuat model terbaik untuk menganalisa dan memprediksi suatu dataset. Dataset *Mushroom Classification* memiliki 8,124 baris dan 23 kolom.

3.4 Preprocessing

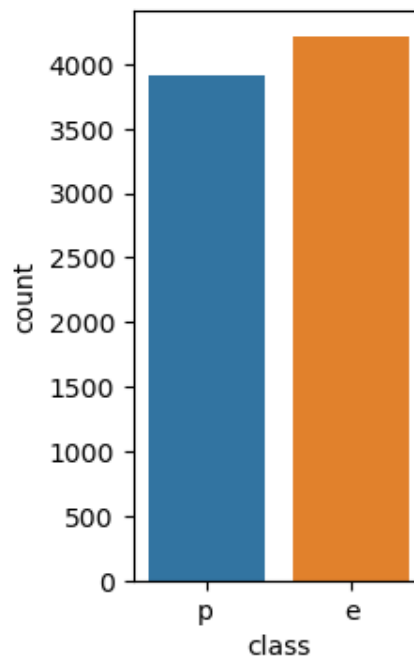
Pada tahap ini, dilakukan beberapa tahapan yang diterapkan untuk memahami data sebelum beralih ke tahapan selanjutnya. Pada bagian ini, dilakukan encoding pada setiap *categorical variabel*, serta dilakukan *reduce dimensionality* pada dataframe.

3.5. Building Model

Pada tahap ini dilakukan train test split pada dataset untuk melakukan evaluasi performa pada dataset yang telah dilakukan preprocessing, membangun model decision tree, memvisualisasikan hasil pembangunan model dan dilakukan prediksi.

IV. HASIL DAN PEMBAHASAN

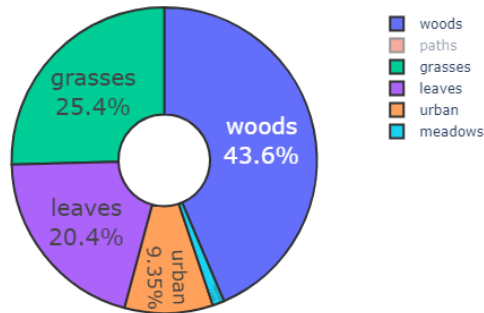
Setelah dilakukan beberapa tahapan klasifikasi didapatkan visualisasi pada *class* sebagai berikut:



Gambar 2. Class

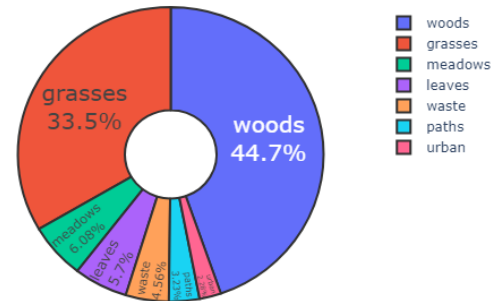
Dimana *edible* (jamur yang dapat dikonsumsi) memiliki jumlah sebanyak 4,208 lebih banyak dari *poisonous* (jamur beracun) yang memiliki jumlah sebanyak 3,916, yang mengartikan bahwa distribusi persebaran pada target variabel seimbang. Pada dataset ini data tidak memiliki *missing value* atau data yang bernilai num atau null. Dataset ini berisikan semua kategorikal kolom, tidak ada numerikal kolom. Untuk mencari tahu dimana lokasi atau habitat yang paling banyak menghasilkan jamur *edible* dan *poisonous* dilakukan visualisasi terhadap class dan habitat menggunakan plotly.express sehingga menampilkan visualisasi sebagai berikut:

Poisonous mushroom according to habitat



Gambar 3. *Habitat Poisonous*

Edible mushroom according to habitat



Gambar 4. *Habitat Edible*

Dari visualisasi tersebut dapat disimpulkan bahwa habitat paling banyak dari jamur *edible* dan *poisonous* adalah terdapat di hutan. Selain hutan, jamur beracun (*poisonous*) dan jamur *edible* (dapat dimakan) juga banyak terdapat di rerumputan.

Encoding dilakukan karena fitur categorical sehingga perlu diubah menjadi numerik. Kemudian, diterapkan PCA. PCA (*Principal Component Analysis*) merupakan salah satu metode reduksi dimensi pada machine learning. Nantinya, PCA akan memilih variabel-variabel yang mampu menjelaskan sebagian besar variabilitas data.

```
array([0.19115608, 0.31696146, 0.43105575, 0.53956497, 0.62080238,
       0.69035599, 0.73374701, 0.76983779, 0.80454079, 0.83552532,
       0.8631652 , 0.88868969, 0.91061711, 0.93120845, 0.9508362 ,
       0.96790118, 0.97830852, 0.98661417, 0.99278388, 0.99693334])
```

Gambar 5. PCA

Berdasarkan hasil PCA yang tertera pada Gambar 5, kami memilih 20 fitur dari total 22 fitur untuk dilakukan *fitting model*, berdasarkan gambar dapat terlihat bahwa 20 fitur dapat menjelaskan sekitar 0.9963% dari 100% keseluruhan data.

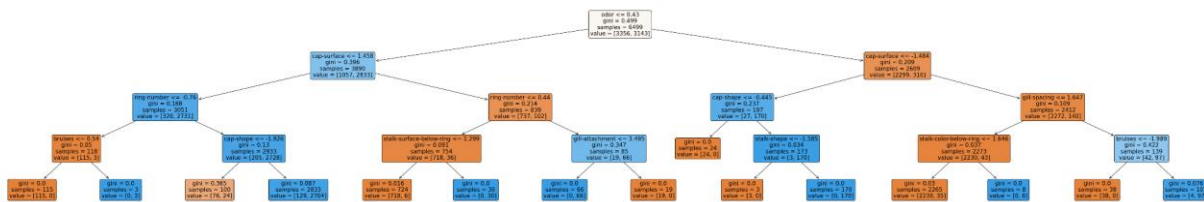
Setelah nilai yang hilang diperhitungkan, data dapat dipecah menjadi bagian validasi-pelatihan (pengujian) dan model dapat dibangun di atas kumpulan data pelatihan. Dataset dibagi menjadi pelatihan dan validasi. StandardScaler dipasang dan diubah pada dataset pelatihan dan transformasi yang sama digunakan pada dataset validasi. StandardScaler menghapus rata-rata dan menskalakan setiap fitur/variabel ke unit variance . Operasi ini dilakukan berdasarkan fitur dengan

cara yang independen.

Shape of Training Data (6499, 20)
Shape of Testing Data (1625, 20)

Gambar 6. Setelah dilakukan train test split

Dataset training dengan observasi 6499 digunakan untuk membangun model dan model yang dibangun divalidasi pada dataset validasi yang memiliki 1625 observasi. Kami melakukan pembangunan model pada dataset training dengan menggunakan criterion “gini” dan max_depth atau level 4, sehingga menghasilkan pohon algoritma seperti berikut:



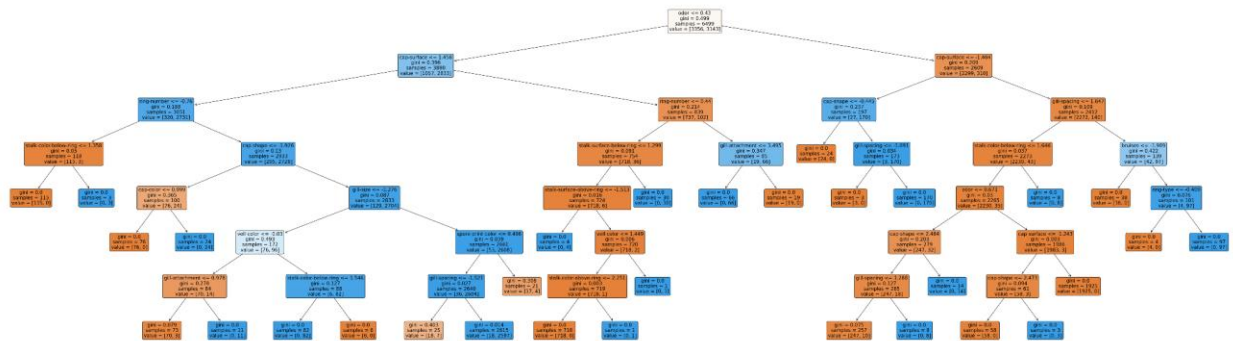
Gambar 7. Decision Tree

Fitur “odor” atau fitur yang menyatakan bau dari jamur ≤ 0.43 dipilih oleh algoritma untuk menjadi akar pada pohon keputusan, penentuan akar pohon ini dipilih berdasarkan nilai gini terbesar. Selanjutnya fitur “odor” dibagi menjadi dua kategori yaitu $\text{cap_surface} < -1.458$ dan $\text{cap_surface} < -1.484$ dimana fitur “cap_surface” merupakan fitur yang menyatakan bentuk permukaan topi jamur. Di node paling bawah, misalnya, Anda memiliki nilai = [4, 97], yang menunjukkan bahwa di antara 101 sampel yang termasuk dalam node tersebut (sampel yang memiliki $\text{odor} < 0.43$, $\text{cap_surface} < -1.484$, $\text{gill_spacing} < 1.647$, dan $\text{bruises} < -1.989$), Oleh karena itu, kemungkinan prediksi = 1 pada node tersebut adalah 97/101. Selanjutnya, kami melakukan hyperparameter tuning untuk mencari kriteria terbaik dalam membangun model, dan melakukan pemodelan ulang berdasarkan hasil dari hyperparameter tuning tersebut.

Best mean accuracy score using grid search:
0.9826127770534552
Best Parameters using grid search:
{‘max_depth’: 7, ‘min_samples_split’: 50}

Gambar 8. hasil hyper parameter tuning

Berdasarkan gambar 8, kami membangun model kembali menggunakan parameter “max_depth” : 7 dan “min_samples_split” : 50



Gambar 9. pembangunan model setelah hyperparameter tuning

V. KESIMPULAN

Laporan ini menjelaskan klasifikasi jamur berdasarkan karakteristiknya berada di kelas yang dapat dikonsumsi atau beracun menggunakan Algoritma Decision Tree. Setelah dilakukan percobaan klasifikasi menggunakan metode Decision Tree, dapat disimpulkan bahwa decision tree merupakan pemodelan yang cukup baik dan sempurna untuk melakukan klasifikasi jamur. Karena dilihat dari evaluasi matriks didapatkan bahwa 849 + 772 mampu memprediksi secara tepat dan memiliki nilai akurasi 99,2% dengan *training set* akurasi 99,3%. Model *Decision Tree Classifier* yang menggunakan indeks gini hanya memiliki perbedaan akurasi model dan akurasi training set yang sangat kecil, sehingga tidak ada tanda-tanda *over fitting*. Untuk mengetahui apakah jamur itu bisa dimakan atau tidak, orang harus terlebih dahulu memeriksa baunya. Jamur yang memiliki bau almond dan adas manis dapat dimakan, sedangkan sisanya, seperti menyengat, busuk, creosote, amis, pedas, dan apek beracun yang berarti tidak bisa dimakan. Untuk jamur yang tidak berbau, ada beberapa atribut yang harus diperiksa seperti spora-print-color, gill-size, gill-spacing, dan populasi.

VI. REFERENSI

- [1] D. L. Hawksworth, *The magnitude of fungal diversity: the 1.5 million species estimate revisited*, vol. 105, Desember 2001.
- [2] I. P. Putra, "Jurnal Ekologi Kesehatan," *KASUS-KASUS KERACUNAN JAMUR LIAR DI INDONESIA*.

VII. LAMPIRAN

Dataset:

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

Code:

<https://colab.research.google.com/drive/11mRSubf2WWW3HZo8Cpa8Zjxl22AYygqP?authuser=2#scrollTo=c858fee3-836e-4a3a-9c94-9d8f9ce71682>

VIII. PEMBAGIAN TUGAS

Robby Hidayah Ramadhan	: membuat codingan
Khoirunnisa	: membuat laporan
Fadia Dilla Sabine	: membuat laporan
Muhammad Rasyid Aditya	: membuat PPT
Heggie Ilham	: membuat PPT