Laporan Akhir Tugas MK. Data Mining (KOM332), Semester Genap 2021/2022

Klasifikasi Jamur Layak Konsumsi Menggunakan *Decision Tree* dan *Random Forest*

Kelompok 7 : Kelas Paralel 2

| Abdul Hakim  G64190078  Departemen Ilmu Komputer  Fakultas Matematika dan Ilmu Pengetahuan Alam  Bogor, Indonesia  *hakimhakim@apps.ipb.ac.id* | Antonius Anre Sianturi  G64190053  Departemen Ilmu Komputer  Fakultas Matematika dan Ilmu Pengetahuan Alam  Bogor, Indonesia  *antoniusanre@apps.ipb.ac.id* | Renny Atikasari  G64190040  Departemen Ilmu Komputer  Fakultas Matematika dan Ilmu Pengetahuan Alam  Bogor, Indonesia  *atikasrrenny@apps.ipb.ac.id* |
| --- | --- | --- |
|  | Ahmad Hadryan Mora  G64190001  Departemen Ilmu Komputer  Fakultas Matematika dan Ilmu Pengetahuan Alam  Bogor, Indonesia  *ryanmora@apps.ipb.ac.id* |  |

***Abstract***

Data mining adalah proses pengumpulan dan pengolahan data yang digunakan untuk mengekstraksi informasi dari database. Penelitian ini dilakukan untuk mengklasifikasikan jenis jamur yang beracun dan jamur yang dapat dikonsumsi. Berdasarkan kasus rata-rata keracunan dikarenakan jamur di Indonesia yang dirangkum oleh jurnal ekologi kesehatan vol.20 No.3 Tahun 2021 menyatakan bahwa telah terjadi 76 kasus rata-rata keracunan jamur di Indonesia akibat pengkonsumsian jamur. Rata-rata sebanyak 550 orang menjadi korban dan dan 9 diantaranya meninggal dunia. Oleh karena itu, tujuan dilakukannya penelitian ini adalah untuk menilai risiko terhadap kehidupan manusia melalui konsumsi jamur. Kumpulan data yang digunakan berisi 8.124 record, di mana 4.208 (52%) contoh dapat dimakan dan 3.916 (48%) beracun. Metode yang digunakan dalam penelitian ini adalah decision tree dan random forest. Hasil daripada penelitian ini menunjukkan bahwa algoritma decision tree dan random forest memiliki akurasi dan presisi yang optimal dalam klasifikasi jamur beracun atau layak konsumsi. Klasifikasi jamur layak konsumsi dan beracun menggunakan algoritma decision tree memberikan hasil akurasi dan juga presisi sebesar **100%** dengan fitur terbaiknya adalah **stalk-surface-above-ring\_k**. Berdasarkan pohon keputusan, permukaan stalk di atas ring yang licin memiliki kemungkinan yang cukup besar jamur tersebut beracun. Sedangkan untuk klasifikasi jamur layak konsumsi dan beracun menggunakan algoritma random forest memberikan hasil akurasi dan juga presisi sebesar **100%** dengan fitur terbaiknya adalah **print-spore-color\_h**. Hal ini menunjukkan bahwa ​​kedua model klasifikasi menghasilkan akurasi dan presisi yang sama besarnya, hanya terdapat perbedaan pada fitur terbaik.

*Kata Kunci : Decision Tree, Jamur Insang, Klasifikasi, Random Forest*

I. PENDAHULUAN

1. *Latar Belakang*

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses Data Mining digunakan untuk mengekstraksi informasi dari database. Data mining selalu berhasil dalam memprediksi data yang digunakan untuk menemukan pola baru dalam database yang besar [1]. Salah satu teknik data mining yaitu klasifikasi menggunakan Decision Tree yang memiliki tugas untuk memprediksi kelas secara akurat yang termasuk dalam sampel data, misalnya sampel data pada jamur.

Jamur merupakan salah satu jenis tumbuhan yang dapat dikonsumsi oleh manusia. Bahkan, jamur juga dapat digunakan sebagai obat-obatan. Tetapi perlu diketahui bahwa tidak semua jamur yang tumbuh di planet ini aman untuk dikonsumsi. Beberapa jenis jamur mengandung racun dan dapat membahayakan nyawa jika dikonsumsi. Meskipun spesies jamur beracun hanya sedikit, dari 70 hingga 80 spesies dari 1,5 juta spesies jamur di dunia, banyak dari jamur mematikan ini memiliki kemiripan yang tidak menguntungkan dengan spesies yang dapat dimakan dan dengan demikian sangat berbahaya.

Kumpulan data jamur yang tersedia dari repositori data UCI dipertimbangkan untuk Pohon Hoeffding. Kumpulan data ini mencakup deskripsi sampel atribut yang sesuai dengan 22 spesies jamur insang dalam famili Agaricus dan Lepiota. Setiap spesies diidentifikasi sebagai pasti dapat dimakan, pasti beracun atau tidak diketahui dapat dimakan dan tidak direkomendasikan. Kumpulan data ini berisi 8.124 record, di mana 4.208 (52%) contoh dapat dimakan dan 3.916 (48%) beracun. Kumpulan data ini dianalisis untuk mengidentifikasi berbagai kombinasi atribut yang signifikan dalam mengklasifikasikan data jamur sebagai beracun atau dapat dimakan.

Berdasarkan kasus rata-rata keracunan dikarenakan jamur di Indonesia yang dirangkum oleh jurnal ekologi kesehatan vol.20 No.3 Tahun 2021 menyatakan bahwa telah terjadi 76 kasus rata-rata keracunan jamur di Indonesia akibat pengkonsumsian jamur. Rata-rata sebanyak 550 orang menjadi korban dan dan 9 diantaranya meninggal dunia [2].

*B. Tujuan*

Tujuan dilakukannya penelitian ini adalah untuk menilai risiko terhadap kehidupan manusia melalui konsumsi jamur. Decision tree dan Random Forest digunakan untuk mengklasifikasikan data dengan 23 atribut, baik jamur yang dapat dimakan maupun yang beracun. Pohon keputusan digunakan dalam penelitian ini sebagai teknik klasifikasi untuk menganalisis kumpulan data jamur.

*C. Ruang Lingkup*

Ruang lingkup dari penelitian ini yaitu :

1. Data set mempunyai 8.124 record
2. Atribut pada dataset berjumlah 23 atribut
3. Algoritma yang digunakan adalah CART
4. Pengambilan keputusan menggunakan Decision Tree dan Random Forest

*D. Manfaat*

Manfaat dari penelitian ini yaitu :

1. Memberikan informasi terkait sebaran jamur insang yang dapat dikonsumsi dan beracun
2. Mampu mengidentifikasi jamur insang yang dapat dikonsumsi dan beracun

TINJAUAN PUSTAKA

*A. Klasifikasi*

Klasifikasi adalah salah satu teknik pada data mining yang memetakan data ke dalam kelompok atau kelas yang telah ditentukan. Klasifikasi merupakan metode *supervised learning* yang membutuhkan data training berlabel untuk menghasilkan sebuah aturan yang mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan [3]. Beberapa teknik klasifikasi yang digunakan adalah *decision tree, rule-based classifier, neural- network, support machine* dan *naïve Bayes classifier*. Setiap teknik menggunakan algoritma pembelajaran untuk mengidentifikasi model yang memberikan hubungan yang paling sesuai antara himpunan atribut dan label kelas dari data input.

*B. Jamur Insang*

Jamur merupakan salah satu jenis tumbuhan yang dijumpai di alam, sehingga sejak dahulu banyak orang yang memanfaatkannya sebagai bahan konsumsi. Salah satu jenis jamur yang sering dijumpai adalah jenis jamur insang dalam famili Agaricus dan Lepiota. jamur insang dalam famili Agaricus dan Lepiota ada yang beracun dan ada yang aman untuk dikonsumsi. jamur insang dalam famili Agaricus dan Lepiota dapat menyebabkan sakit terhadap seseorang yang mengkonsumsinya, bahkan juga dapat menyebabkan kematian.

*C. Pra-proses Data*

Data Pre-Processing merupakan transformasi data mentah yang diolah menjadi kumpulan data yang terbentuk dengan baik, sehingga analitik data mining dapat diterapkan. Data mentah seringkali tidak lengkap dan memiliki format yang tidak konsisten. Kecukupan dan ketidakcukupan persiapan data memiliki korelasi langsung dengan keberhasilan setiap proyek yang melibatkan analisis data. Pre-processing melibatkan validasi data dan imputasi data. Tujuan dari validasi data adalah untuk menilai apakah data yang bersangkutan sudah lengkap dan akurat.

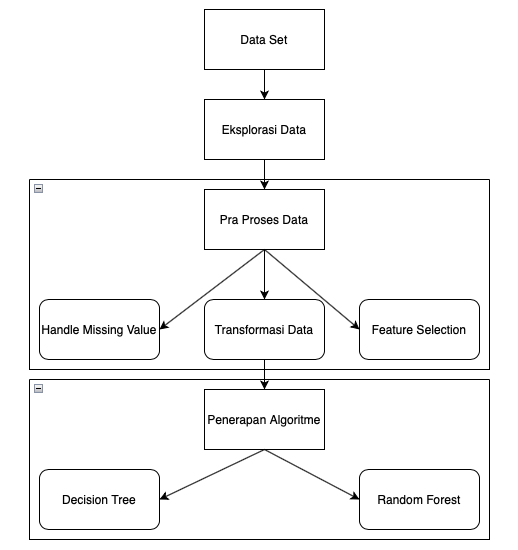
III. METODOLOGI

Metode yang digunakan pada penelitian ini yaitu klasifikasi yang menggunakan teknik Decision Tree dan Random Forest. Pada penelitian ini data jamur akan dibagi menjadi 2, data latih dan data uji dengan persentase 70% dan 30%. Pada data latih akan dibentuk model klasifikasi terarah menggunakan decision tree dan random forest. Adapun ukuran kebaikan yang digunakan adalah persentase akurasi dan presisi dari penerapan model terhadap data uji. Presisi digunakan sebagai penentu kebaikan model karena pada penelitian kali ini kita tidak ingin jamur yang beracun diklasifikasikan sebagai jamur yang layak konsumsi. Dengan kata lain, kita ingin minimalisasi *false* layak konsumsi.

*A. Data*

Kumpulan data jamur yang tersedia dari repositori data UCI. Kumpulan data ini mencakup deskripsi sampel yang memiliki 23 atribut yang sesuai dengan spesies jamur insang dalam famili Agaricus dan Lepiota. Seluruh atribut dari data merupakan data kategorik nominal yang disimpan dalam tipe data string. Setiap spesies diidentifikasi sebagai pasti dapat dimakan, pasti beracun atau tidak diketahui dapat dimakan dan tidak direkomendasikan. Kasus jamur yang tidak direkomendasikan digabungkan dengan yang beracun. Kumpulan data ini berisi 8.124 record, di mana 4.208 (52%) contoh dapat dimakan dan 3.916 (48%) beracun. Kumpulan data ini dianalisis untuk mengidentifikasi berbagai kombinasi atribut yang signifikan dalam mengklasifikasikan data jamur sebagai beracun atau dapat dimakan.

*B. Tahapan Kegiatan*

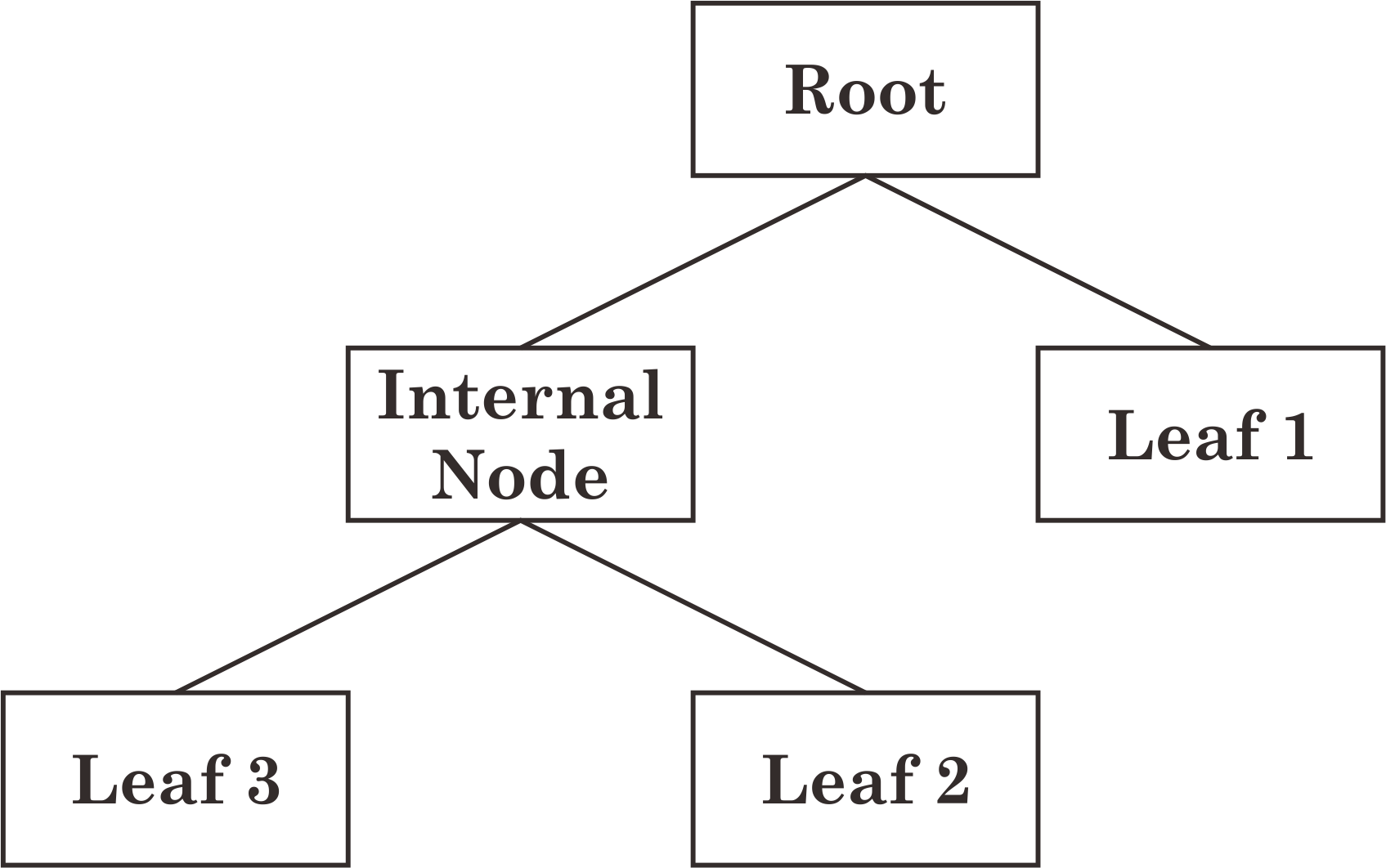
**

*Gambar 1. Tahapan Penelitian*

*C. Lingkungan Pengembangan*

*C.1 Decision Tree*

Pohon keputusan merupakan salah satu metode klasifikasi yang popular karena dapat dengan mudah diinterpretasi oleh manusia. Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan [4].

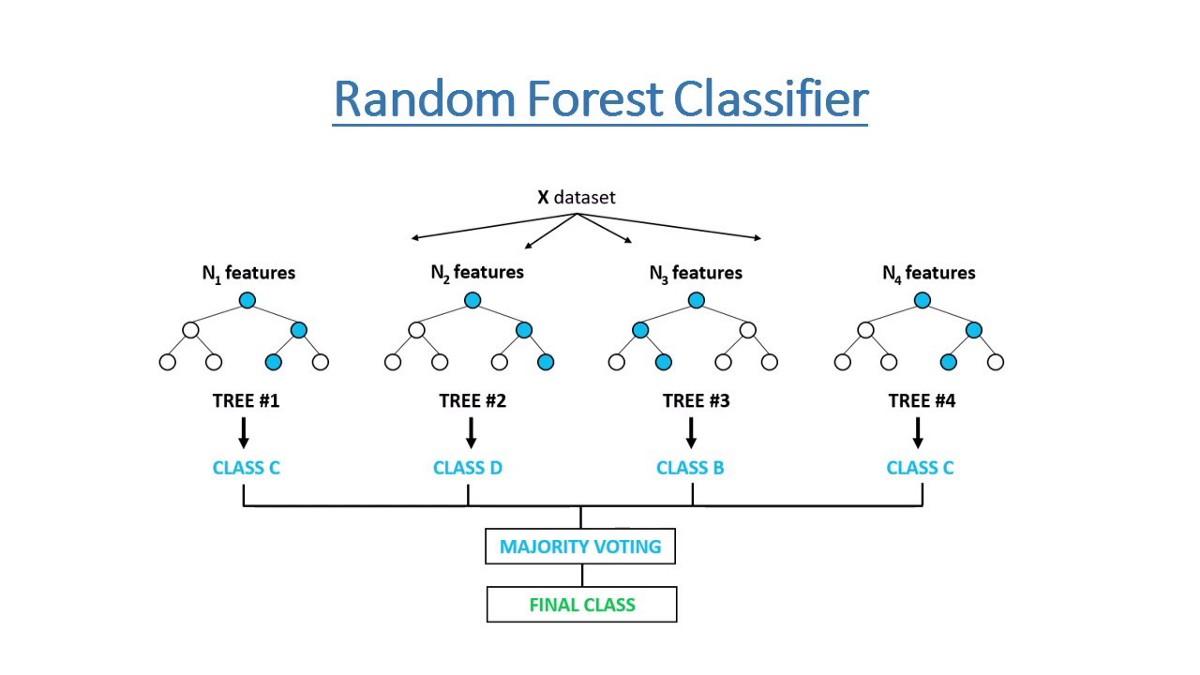


*Gambar 2. Decision Tree*

Pohon keputusan memiliki node pohon yang merepresentasikan atribut yang telah diuji dan setiap cabangnya merupakan suatu pembagian hasil uji serta node daun (*leaf*) merepresentasikan kelompok kelas tertentu [5]. Level node teratas dari sebuah pohon keputusan adalah node akar (*root*) yang biasanya berupa atribut yang memiliki pengaruh paling besar pada suatu kelas tertentu. Konsep dasar dari pohon keputusan adalah mengubah data menjadi model pohon keputusan, kemudian mengubah model pohon menjadi rule dan menyederhanakan rule. Data dalam pohon keputusan dinyatakan dalam bentuk tabel dengan atribut dan record.

*C.2 Random Forest*

Metode *Random Forest*  merupakan salah satu metode dalam *Decision Tree*. *Random forest*  adalah kombinasi dari masing – masing *tree* yang baik kemudian dikombinasikan ke dalam satu model. *Random Forest* bergantung pada sebuah nilai vector random dengan distribusi yang sama pada semua pohon yang masing masing *decision tree* memiliki kedalaman yang maksimal. *Random forest* adalah *classifier* yang terdiri dari *classifier* yang berbentuk pohon {h(**x**, θ k ), k = 1, . . .} dimana θ*k* adalah random vektor yang didistribusikan secara independen dan masing masing tree pada sebuah unit kan memilih class yang paling popular pada input x. Berikut ini ilustrasi dari *random forest* [6].

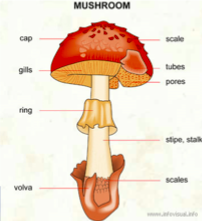


*Gambar 3. Random Forest*

IV. HASIL DAN PEMBAHASAN

*A. Eksplorasi Data*

Pada penelitian ini, yang pertama dilakukan adalah eksplorasi data. Hal ini dilakukan untuk melihat apa saja yang menjadi atribut data, statistik deskriptif data, perbandingan jumlah class, dan korelasi antara atribut.



*Gambar 4. Struktur jamur*

Pada struktur jamur yang bisa dilihat pada *Gambar 1. Struktur jamur*, maka didapatkan data atribut yaitu :

Attribute Information: (classes: edible=e, poisonous=p)

* cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
* cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
* Cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y
* bruises: bruises=t,no=f
* odor: almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m,none=n,pungent=p,spicy=s
* gill-attachment: attached=a,descending=d,free=f,notched=n
* gill-spacing: close=c,crowded=w,distant=d
* gill-size: broad=b,narrow=n
* gill-color:black=k, brown=n, buff=b, chocolate=h, gray=g, green=r,orange=o,pink=p,purple=u,red=e,white=w,yellow=y
* stalk-shape: enlarging=e,tapering=t
* stalk-root: bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r,missing=?
* stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
* stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
* stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
* stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
* veil-type: partial=p,universal=u
* veil-color: brown=n,orange=o,white=w,yellow=y
* ring-number: none=n,one=o,two=t
* ring-type: cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing=s,zone=z
* spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r,orange=o,purple=u,white=w,yellow=y
* population: abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y
* habitat: grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d

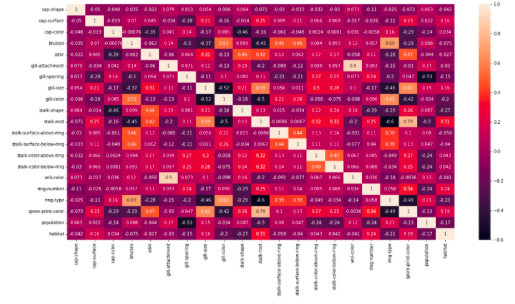
Berikut statistik deskriptif dari data untuk setiap atributnya. Dikarenakan data bertipe kategorik, maka beberapa statistik yang ditampilkan adalah jumlah data, banyaknya nilai/kategori, modus, dan jumlah modus dari setiap atributnya.

| **Atribut** | **Jumlah** | **Jumlah Kategori** | **Modus** | **Jumlah Modus** |
| --- | --- | --- | --- | --- |
| **class** | 8124 | 2 | e | 4208 |
| **cap-shape** | 8124 | 6 | x | 3656 |
| **cap-surface** | 8124 | 4 | y | 3244 |
| **cap-color** | 8124 | 10 | n | 2284 |
| **bruises** | 8124 | 2 | f | 4748 |
| **odor** | 8124 | 9 | n | 3528 |
| **gill-attachment** | 8124 | 2 | f | 7914 |
| **gill-spacing** | 8124 | 2 | c | 6812 |
| **gill-size** | 8124 | 2 | b | 5612 |
| **gill-color** | 8124 | 12 | b | 1728 |
| **stalk-shape** | 8124 | 2 | t | 4608 |
| **stalk-root** | 5644 | 4 | b | 3776 |
| **stalk-surface-above-ring** | 8124 | 4 | s | 5176 |
| **stalk-surface-below-ring** | 8124 | 4 | s | 4936 |
| **stalk-color-above-ring** | 8124 | 9 | w | 4464 |
| **stalk-color-below-ring** | 8124 | 9 | w | 4384 |
| **veil-type** | 8124 | 1 | p | 8124 |
| **veil-color** | 8124 | 4 | w | 7924 |
| **ring-number** | 8124 | 3 | o | 7488 |
| **ring-type** | 8124 | 5 | p | 3968 |
| **spore-print-color** | 8124 | 9 | w | 2388 |
| **population** | 8124 | 6 | v | 4040 |
| **habitat** | 8124 | 7 | d | 3148 |

*Tabel 1. Statistik deskriptif atribut*

Selanjutnya, pada eksplorasi data juga didapatkan perbandingan jumlah class yang ditemukan hampir setara dan juga korelasi antar atribut yang bisa dilihat pada gambar berikut : 

*Gambar 5. Perbandingan jumlah class*

*Gambar 6. Korelasi antara atribut*

Berdasarkan heat map atau korelasi atribut, dapat diketahui bahwa gill-attachment dan veil-color memiliki korelasi cukup besar. Artinya, besar kemungkinan nilai gill-attachment bergantung pada veil-color dan juga sebaliknya.

*B. Pra-Proses Data*

Data Pre-Processing merupakan transformasi data mentah yang diolah menjadi kumpulan data yang terbentuk dengan baik, sehingga analitik data mining dapat diterapkan. Data mentah seringkali tidak lengkap dan memiliki format yang tidak konsisten. Kecukupan dan ketidakcukupan persiapan data memiliki korelasi langsung dengan keberhasilan setiap proyek yang melibatkan analisis data. Pre-processing melibatkan validasi data dan imputasi data. Tujuan dari validasi data adalah untuk menilai apakah data yang bersangkutan sudah lengkap dan akurat. Ada beberapa hal terkait pra-proses data yang dilakukan dalam penelitian kali ini, yaitu meliputi : *handle missing value*, transformasi data, *feature selection*, dan pembagian data.

Hal pertama yang dilakukan pada pra-proses data yaitu *handle missing value.* Missing value pada data jamur terdapat pada bagian stalk-root sebanyak 2480 dari 8124 record data. Jumlah record dari data yang hilang sebesar 30.5%. Berdasarkan hal tersebut, maka dapat disimpulkan bahwa peneliti tidak bisa menghilangkan record yang memiliki missing value pada bagian stalk-root. Oleh karena itu, penanganan yang dilakukan adalah menghilangkan stalk-root dari atribut yang akan digunakan untuk model.

| **Atribut** | **Missing Value** | **Atribut** | **Missing Value** |
| --- | --- | --- | --- |
| cap-shape | 0 | stalk-root | 2480 |
| cap-surface | 0 | veil-type | 0 |
| cap-color | 0 | veil-color | 0 |
| bruises | 0 | ring-number | 0 |
| odor | 0 | ring-type | 0 |
| gill-attach | 0 | spore-print | 0 |
| gill-spacing | 0 | population | 0 |
| gill-size | 0 | habitat | 0 |
| stalk-shape | 0 | gill-color | 0 |
| stalk-color-above-ring | 0 | stalk-color-below-ring | 0 |
| stalk-surface-above-ring | 0 | stalk-surface-below-ring | 0 |

*Tabel 2. Missing value atribut*

Selanjutnya, dilakukan transformasi data. Hal pertama yang dilakukan pada transformasi data adalah penghapusan beberapa atribut, yaitu :

* Stalk-root dihapus dari calon pembentuk model karena merupakan missing value.
* Veil-type dihapus dari calon pembentuk model dikarenakan jenis pada atribut tersebut hanya terdapat satu sehingga tidak dapat digunakan sebagai pembeda.
* Class dihapus dari calon pembentuk model karena merupakan atribut yang menjadi target.

Kemudian, pada transformasi data dilakukan pengkodean untuk setiap atribut. Atribut data yang dimiliki berupa data nominal dengan tipe data string. Maka dari itu, untuk melakukan proses seleksi fitur, data perlu diubah bentuk menjadi tipe data integer menggunakan label encoder.

| **index** | **0** | **1** | **2** | **3** | **4** |
| --- | --- | --- | --- | --- | --- |
| class | 1 | 0 | 0 | 1 | 0 |
| cap-shape | 5 | 5 | 0 | 5 | 5 |
| cap-surface | 2 | 2 | 2 | 3 | 2 |
| cap-color | 4 | 9 | 8 | 8 | 3 |
| bruises | 1 | 1 | 1 | 1 | 0 |
| odor | 6 | 0 | 3 | 6 | 5 |
| gill-attachment | 1 | 1 | 1 | 1 | 1 |
| gill-spacing | 0 | 0 | 0 | 0 | 1 |
| gill-size | 1 | 0 | 0 | 1 | 0 |
| gill-color | 4 | 4 | 5 | 5 | 4 |
| stalk-shape | 0 | 0 | 0 | 0 | 1 |
| stalk-root | 2 | 1 | 1 | 2 | 2 |
| stalk-surface-above-ring | 2 | 2 | 2 | 2 | 2 |
| stalk-surface-below-ring | 2 | 2 | 2 | 2 | 2 |
| stalk-color-above-ring | 7 | 7 | 7 | 7 | 7 |
| stalk-color-below-ring | 7 | 7 | 7 | 7 | 7 |
| veil-type | 0 | 0 | 0 | 0 | 0 |
| veil-color | 2 | 2 | 2 | 2 | 2 |
| ring-number | 1 | 1 | 1 | 1 | 1 |
| ring-type | 4 | 4 | 4 | 4 | 0 |

*Tabel 3. Label Encode atribut*

Pra-proses data yang berikutnya yaitu *feature selection.* Fitur atau atribut yang akan digunakan diseleksi berdasarkan nilai dari atribut terhadap kelas. Nilai digunakan karena dapat menghitung dependensi dari setiap atribut terhadap *class* yang menjadi target. Banyaknya jumlah atribut yang akan dipilih adalah 10 menggunakan SelectKBest. Adapun fitur yang didapatkan sebagai berikut dengan score nya :

| **No** | **Atribut** | **Score** |
| --- | --- | --- |
| **1** | gill-color | 5957.764469 |
| **2** | ring-type | 1950.610146 |
| **3** | gill-size | 1636.606833 |
| **4** | bruises | 1194.277352 |
| **5** | gill-spacing | 826.795274 |
| **6** | habitat | 751.309489 |
| **7** | spore-print-color | 379.132729 |
| **8** | population | 311.766736 |
| **9** | stalk-surface-above-ring | 222.982400 |
| **10** | cap-surface | 214.068544 |

*Tabel 4. Nilai score chi selectKBest*

Setelah didapatkan fitur yang akan digunakan untuk membentuk model, tahapan berikutnya adalah membentuk peubah dummy berdasarkan fitur yang terseleksi. Seluruh kategori dalam sebuah atribut akan menjadi suatu atribut sendiri yang memiliki nilai 0 atau 1. Nilai 0 memiliki arti atribut tersebut tidak terdapat pada satu record yang diamati dan nilai 1 memiliki arti atribut tersebut terdapat pada satu record yang diamati. Peubah dummy membuat setiap atribut hanya terdapat 2 kategori (*True or False)* dengan harapan bisa meningkatkan performa model yang hanya perlu memecah setiap atribut menjadi 2. Adapun peubah dummy yang akan dihasilkan sebanyak 53 dan akan digunakan sebagai calon pembentuk model.

Tahap terakhir dari pra-proses data adalah pembagian data latih dan uji. Data uji akan digunakan sebagai validasi kebaikan model. Melalui data uji, akan dilihat akurasi dan juga presisi dari model yang dibentuk dari data latih. Proporsi pembagian data latih dan uji adalah 70% dan 30% sehingga didapatkan total data latih sebanyak 5686 dan data uji sebanyak 2438.

*C. Penerapan Algoritma*

C.1 Decision tree

Model klasifikasi jamur layak konsumsi dan beracun akan dibentuk menggunakan algoritma pohon keputusan berdasarkan data latih. Pembentukan model pada penelitian ini menggunakan algoritma CART (Classification and Regression Tree) yang tersedia pada package sklearn dalam bahasa pemrograman python. Dalam pembentukan model, dilakukan hyperparameter tuning menggunakan Grid Search yang juga merupakan sebuah fungsi tersedia di python pada package sklearn. Adapun metode yang dilakukan Grid Search dalam mencari parameter terbaik pada penelitian ini adalah melakukan cross-validation dengan 5 fold. Setiap variasi parameter akan dilakukan pembentukan model, lalu akan dicari nilai rata-rata akurasi terbesar sebagai parameter terbaik. Berikut parameter pembentuk decision tree yang diikutsertakan dalam hyperparameter tuning :

* Splitter : best, random
* Criterion : gini, entropy
* Min\_samples\_split : 2, 3, 4, 5
* Max\_depth : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
* Class\_weight : balanced, none

Setelah dilakukan pembentukan model terhadap 384 variasi parameter, didapatkan model dengan parameter terbaik sebagai berikut :

* splitter = best
* criterion = gini
* min\_samples\_split = 2
* max\_depth = 9
* class\_weight : balanced

C.2 Random forest

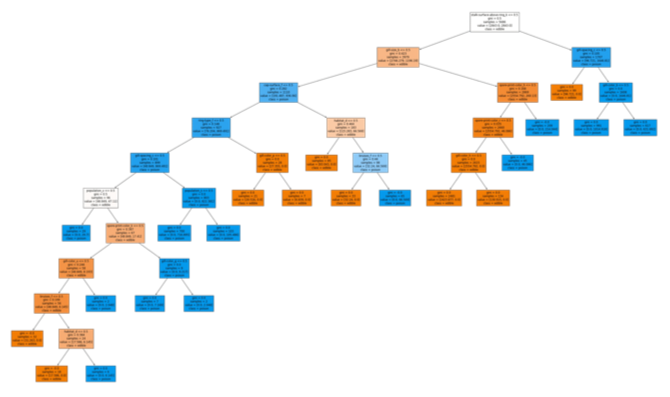
Model klasifikasi jamur layak konsumsi dan beracun akan dibentuk menggunakan algoritma random forest berdasarkan data latih. Pembentukan model pada penelitian ini menggunakan algoritma CART (Classification and Regression Tree) yang tersedia pada package sklearn dalam bahasa pemrograman python. Dalam pembentukan model, dilakukan hyperparameter tuning menggunakan Grid Search yang juga merupakan sebuah fungsi tersedia di python pada package sklearn. Hyperparameter tuning dilakukan untuk mendapatkan parameter paling optimal dari model yang dibentuk. Berikut parameter yang digunakan sebagai pembentuk random forest yang diikutsertakan dalam hyperparameter tuning :

* 'criterion' :['gini', 'entropy'],
* 'min\_samples\_split':[2, 3, 4, 5],
* 'max\_depth':[1, 2, 3, 4, 5, 6, 7, 8],
* 'class\_weight':('balanced', None)

Model terbaik yang didapatkan dari parameter :

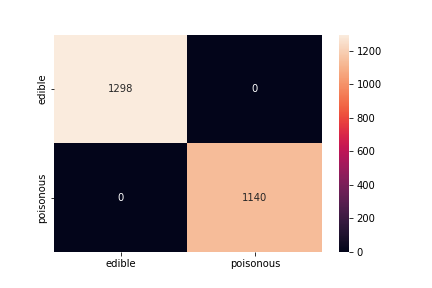
* criterion = gini
* min\_samples\_split = 2
* max\_depth = 7
* class\_weight : balanced

*D. Analisis Hasil Decision Tree*



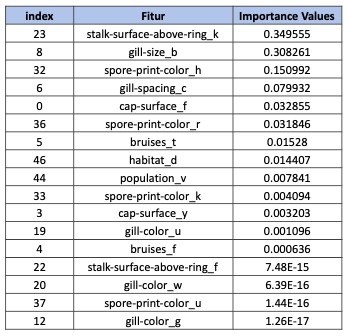
*Gambar 7. Hasil Decision Tree*

Model decision tree dibentuk dan didapatkan gambar pohon keputusan seperti yang tertera pada gambar di atas. Pada bagian akar atau *root* terdapat atribut stalk\_surface\_above\_ring\_k.



*Gambar 8. Confusion Matrix Decision Tree*

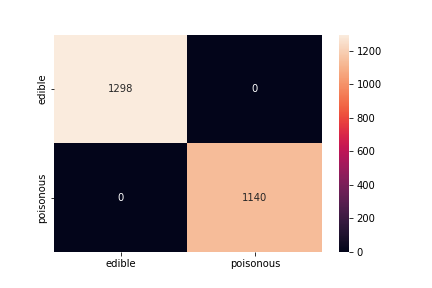
Setelah dilakukan penerapan algoritma decision tree didapatkan hasil yang tertera pada confusion matrix di atas, dapat dilihat bahwa dari model klasifikasi decision tree sebesar 100% begitu juga dengan presisi yang menjadi tolak ukur kebaikan model. Selain itu, dari hasil decision tree di dapatkan juga feature importance yang bisa menjadi tolak ukur kepentingan suatu atribut terhadap beracun atau layak konsumsinya jamur tertentu.



*Tabel 5. Feature Importance Decision Tree*

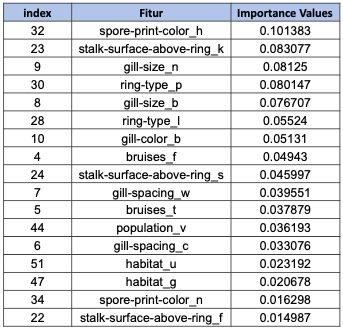
Melalui model decision tree bisa dilihat fitur-fitur apa saja yang memiliki peran penting dalam menentukan keputusan. Fitur yang baik adalah fitur yang bisa menjadi pembeda kelas jamur layak konsumsi dan beracun. Atribut stalk\_surface\_above\_ring merupakan atribut terbaik yang juga menjadi root dari pohon keputusan. Berdasarkan tree yang dihasilkan, jamur yang memiliki permukaan stalk di atas ring berbentuk silky mempunyai kemungkinan yang cukup besar jamur tersebut diklasifikasikan sebagai beracun.

*E. Analisis Hasil Random Forest*



*Gambar 9. Confusion Matrix Random Forest*

Setelah dilakukan penerapan algoritma *random forest* didapatkan hasil yang tertera pada confusion matrix di atas, dapat dilihat bahwa dari model klasifikasi *random forest* sebesar 100% begitu juga dengan presisi yang menjadi tolak ukur kebaikan model. Selain itu, dari hasil *random forest* di dapatkan juga *feature importance* yang bisa menjadi tolak ukur kepentingan suatu atribut terhadap beracun atau layak konsumsinya jamur tertentu.

**

*Tabel 6. Fitur Importance Random Forest*

Pada algoritma *random forest* didapatkan fitur-fitur terbaik seperti yang tertera pada tabel. Fitur terbaik adalah spore\_print\_color\_h yang merupakan warna cetak dari spora pada jamur berwarna coklat. Berbeda dengan *decision tree,* tidak dapat menentukan kemungkinan jamur layak konsumsi atau beracun. Namun, tetap menyatakan bahwa warna spora coklat bisa membedakan jamur layak konsumsi dan beracun dengan baik.

V. KESIMPULAN DAN SARAN

1. *Kesimpulan*

Dalam penelitian ini kami menggunakan algoritma klasifikasi decision tree dan random forest. Tujuan kami adalah untuk menilai risiko terhadap kehidupan manusia melalui jamur yang kami konsumsi apakah termasuk jenis jamur beracun atau jamur yang layak dan aman untuk dikonsumsi. Penelitian ini berkonsentrasi pada dataset jamur. Hasil penelitian ini menunjukkan akurasi dan presisi untuk kinerja yang diukur dari algoritma klasifikasi menggunakan decision tree dan random forest. Hasil yang didapatkan dari penelitian jamur beracun atau layak konsumsi dengan decision tree dan random forest menunjukkan hasil yang sama-sama optimal yaitu 100%. Selain itu, berdasarkan kebaikan fiturnya dalam membentuk algoritma, atribut permukaan stalk di atas ring berbentuk silky dan warna spora coklat merupakan pembeda yang baik antara jamur layak konsumsi dan beracun.

1. *Saran*

* Berdasarkan penelitian yang dilakukan menggunakan decision tree dan random forest, maka didapatkan hasil akurasi yang sama-sama optimal yaitu 100%. Hasil yang didapatkan bisa menimbulkan beberapa kecurigaan sehingga sebaiknya dilakukan pengecekan ulang dengan cara membagi lagi partisi data menjadi 10 kemudian dilakukan train dan test untuk setiap partisi data. Setelah itu, untuk mendapatkan hasil yang sebenarnya maka setiap hasil dari pengecekan 10 partisi data diambil nilai rata-ratanya.
* Penarikan kesimpulan berdasarkan hasil penelitian yang mencapai akurasi 100% menimbulkan asumsi bahwa data pattern yang digunakan mudah dan tidak variatif. Kedepannya, mungkin bisa menggunakan dataset yang lebih bervariasi dengan pattern yang lebih sulit.
* Untuk metode penelitian menggunakan random forest dianggap tidak perlu. Dikarenakan random forest merupakan turunan dari decision tree dan akurasi yang didapatkan dari penelitian menggunakan decision tree sudah mendapatkan akurasi yang sangat optimal yaitu 100%. Sehingga penelitian menggunakan random forest dianggap tidak perlu dilakukan lagi.

VI. DAFTAR PUSTAKA

[1] Han J., Kamber M. and Pei J. **(2012).** Data Mining: Concepts and Techniques, Morgan Kaufmann.

[2] [Vol 20 No 3 (2021): JURNAL EKOLOGI KESEHATAN VOLUME 20 NOMOR 3 TAHUN 2021](https://ejournal2.litbang.kemkes.go.id/index.php/jek/issue/view/369)

[3] Dunham, M.H. (2003). Data Mining: Introductory and Advanced Topics, Pearson Education Inc.

[4] Berry, Michael J.A. & Gordon S. Linoff. (2004). Data Mining Techniques for Marketing, Sales and Customer Relationship Management, Second Edition, Wiley Publishing, Inc.

[5] Han, J. & M. Kamber. (2006). Data Mining Concept and Techniques, Morgan Kaufmann Publishers, San Fransisco.

[6] Ho, Tin Kam (1995). [Random Decision Forests](http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf) (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.