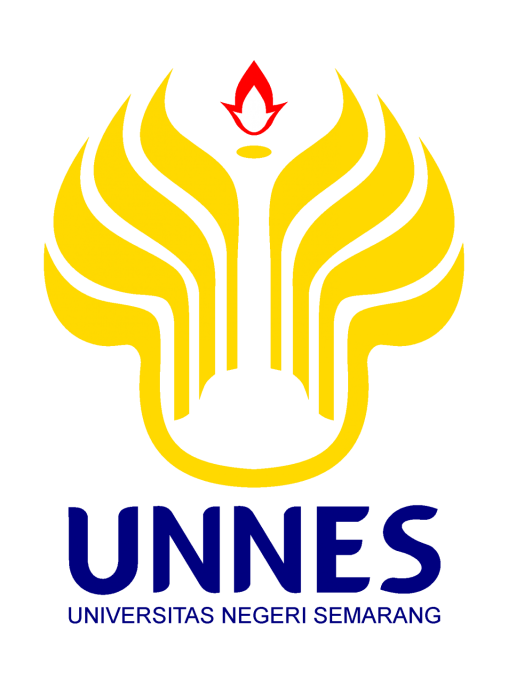
**MAKALAH STATISTIKA MULTIVARIAT**

**KLASIFIKASI DENGAN METODE DECISION TREE (POHON KEPUTUSAN)**

**(SUPERVISED LEARNING)**



**Di Susun Oleh :**

**Reza Dilla Saputri (4112317010)**

**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SEMARANG**

**2019**

# **KATA PENGANTAR**

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat-Nya sehingga makalah ini dapat tersusun hingga selesai .Tidak lupa kami juga mengucapkan banyak terimakasih atas bantuan dari pihak yang telah berkontribusi dengan memberikan sumbangan baik materi maupun pikirannya.

Dan harapan kami semoga makalah ini dapat menambah pengetahuan dan pengalaman bagi para pembaca, Untuk ke depannya dapat memperbaiki bentuk maupun menambah isi makalah agar menjadi lebih baik lagi.

Karena keterbatasan pengetahuan maupun pengalaman kami, Kami yakin masih banyak kekurangan dalam makalah ini, Oleh karena itu kami sangat mengharapkan saran dan kritik yang membangun dari pembaca demi kesempurnaan makalah ini.

Semarang, 15 Juni 2019

Penyusun

**DAFTAR ISI**

[Kata Pengantar 2](#_Toc12742153)

[**BAB I** 4](#_Toc12742154)

[**PENDAHULUAN** 4](#_Toc12742155)

[Latar Belakang 4](#_Toc12742156)

[Rumusan Masalah 4](#_Toc12742157)

[Tujuan 4](#_Toc12742158)

[**BAB II** 5](#_Toc12742159)

[**PEMBAHASAN** 5](#_Toc12742160)

[**1. Penjelasan Decision Tree** 5](#_Toc12742161)

[Manfaat Decision Tree (pohon keputusan) 5](#_Toc12742163)

[Kelebihan Decision Tree (Pohon Keputusan) 6](#_Toc12742164)

[Kekurangan Decision Tree (Pohon Keputusan) 7](#_Toc12742165)

[Model Decision Tree (Pohon Keputusan) 7](#_Toc12742166)

[Tahap tahap proses klasifikasi 8](#_Toc12742167)

[**2.** **Contoh Penerapan Analisis Decision Tree** 9](#_Toc12742168)

[**BAB III** 16](#_Toc12742192)

[**PENUTUP** 16](#_Toc12742193)

[Kesimpulan 16](#_Toc12742194)

[Saran 16](#_Toc12742196)

[**DAFTAR PUSTAKA** 17](#_Toc12742198)

BAB I

PENDAHULUAN

## Latar Belakang

Supervised learningadalah sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengkelompokan suatu data ke data yang sudah ada.Tugas supervised learning adalah untuk membangun sebuah classifier dengan memberikan sekumpulan contoh traiing yang sudah diklasifikasikan .

Supervised learning tidak hanya mempelajari classifier,tetapi juga mempelajari fungsi yang dapat memprediksi suatu nlai numeric.Contoh: ketika diberi fotot seseorang ,kita ingn memprediksi umur,tinggi,dan berat orang yang ada pada foto tersebut.

Ada banyak algoritma pembelajaran yang dikembangkan dari supervised learning.Algoritma algoritma tersebut adalah decision trees (pohon keputusan) ,artificial neural networks,support vector machine.Pada makalah ini akan dibahas mengenai salah satu metode supervised learning dengan decision tree (pohon keputusan)

## Rumusan Masalah

1. Apa yang dimaksud analisis random forest dengan decision tree ?
2. Bagaimana contoh penerapan dari analisis random forest di R Studio ?

## Tujuan

1. Untuk mengetahui tentang apa itu analisis random forest dengan decision tree
2. Untuk mengetahui contoh penerapan analisis random forest di R Studio

BAB II

PEMBAHASAN

1. Penjelasan Decision Tree

Decision tree adalah salah satu metode klasifikasi yang paling populer, karena mudah untuk diinterpretasi oleh manusia. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

Konsep dari pohon keputusan adalah mengubah data menjadi decision tree dan aturan-aturan keputusan. Manfaat utama dari penggunaan decision tree adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Nama lain dari decision tree adalah CART (Classification and Regression Tree). Dimana metode ini merupakan gabungan dari dua jenis pohon, yaitu classification tree dan juga regression tree. Untuk memudahkan, berikut ilustrasi dari keduanya.

Decision tree juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. *Decision tree* memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Dalam beberapa aplikasi, akurasi dari sebuah klasifikasi atau prediksi adalah satu-satunya hal yang ditonjolkan dalam metode ini, misalnya sebuah perusahaan *direct mail* membuat sebuah model yang akurat untuk memprediksi anggota mana yang berpotensi untuk merespon permintaan, tanpa memperhatikan bagaimana atau mengapa model tersebut bekerja.

### ****Manfaat Decision Tree (pohon keputusan)****

Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi  
antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memadukan antara  
eksplorasi data dan pemodelan, sehingga  sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika  
dijadikan sebagai model akhir dari beberapa teknik lain. Sering terjadi tawar menawar antara keakuratan  
model dengan transparansi model. Dalam beberapa aplikasi, akurasi dari sebuah klasifikasi atau prediksi adalah satu-satunya hal yang ditonjolkan, misalnya sebuah perusahaan direct mail membuat sebuah model yang akurat untuk  
memprediksi anggota mana yang berpotensi untuk merespon permintaan, tanpa memperhatikan bagaimana atau mengapa model tersebut bekerja.

### ****Kelebihan Decision Tree (Pohon Keputusan)****

Kelebihan dari metode pohon keputusan adalah:

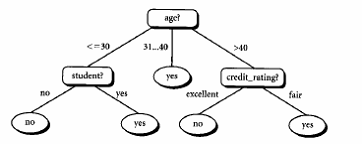
* Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
* Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
* Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama. Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional
* Dalam analisis multivariat, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasikan baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan criteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

### ****Kekurangan Decision Tree (Pohon Keputusan)****

* Terjadi overlap terutama ketika kelas-kelas dan criteria yang digunakan jumlahnya sangat banyak. Hal tersebut juga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
* Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.
* Kesulitan dalam mendesain pohon keputusan yang optimal.
* Hasil kualitas keputusan yang didapatkan dari metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut didesain.

### ****Model Decision Tree (Pohon Keputusan)****

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Contoh dari pohon keputusan dapat dilihat di Gambar berikut ini.

[](https://fairuzelsaid.files.wordpress.com/2009/11/image002.gif)

Model Pohon Keputusan (Pramudiono,2008)

Disini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Contoh di Gambar 1 adalah identifikasi pembeli komputer,dari pohon keputusan tersebut diketahui bahwa salah satu kelompok yang potensial membeli komputer adalah orang yang berusia di bawah 30 tahun dan juga pelajar. Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklasifikasikan record yang belum ada kelasnya. Dimulai dari node root, menggunakan tes terhadap atribut dari record yang belum ada kelasnya tersebut lalu mengikuti cabang yang sesuai dengan hasil dari tes tersebut, yang akan membawa kepada internal node (node yang memiliki satu cabang masuk dan dua atau lebih cabang yang keluar), dengan cara harus melakukan tes lagi terhadap atribut atau node daun. Record yang kelasnya tidak diketahui kemudian diberikan kelas yang sesuai dengan kelas yang ada pada node daun. Pada pohon keputusan setiap simpul daun menandai label kelas. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (tree) kemudian mengubah model pohon tersebut menjadi aturan (rule).

* **Tahap tahap proses klasifikasi**

Tahap awal yang dilakukan sebelum proses klasifikasi adalah melakukan preprocessing data

terdapat beberapa langkah  preprocessing data,yaitu :

a.Data Integrasi

Data Integrasi merupakan penggabungan data dari beberapa sumber. Dalam metode klasifikasi data integrase dilakukan dengan mengumpulkan beberapa data yang berasal dari beberapa departemen di sebuah perusahaan untuk mendapatkan satu database yang akan digunakan untuk pengolahan pada metode klasifikasi.

b.Data Cleaning

Data cleaning merupakan suatu pemrosesan terhadap data untuk menghilangkan noise dan penanganan terhadap missing value pada suatu record.

c.Data Diskretisasi

Data diskretisasi merupakan proses membuat kelas kelas data (range) untuk parameter dalam melakukan transformasi untuk tahapan berikutnya.

d.Transformasi Data

Pada data dapat dilakukan generalisasi menjadi data dengan level yang lebih tinggi. Misalnya dengan melakukan diskretisasi terhadap atribut dengan nilai kontinyu. Pembelajaran terhadap data hasil generalisasi dapat mengurangi kompleksitas pembelajaran yang harus dilakukan karena ukuran data yang harus diproses lebih kecil.

e.Data Reduksi

 Data reduksi merupakan proses mengurangi data atau variable yang tidak memiliki hubungan atau korelasi dengan tujuan penelitian. Dalam menggunakan metode klasifikasi, data reduksi dilakukan dengan melakukan uji independensi

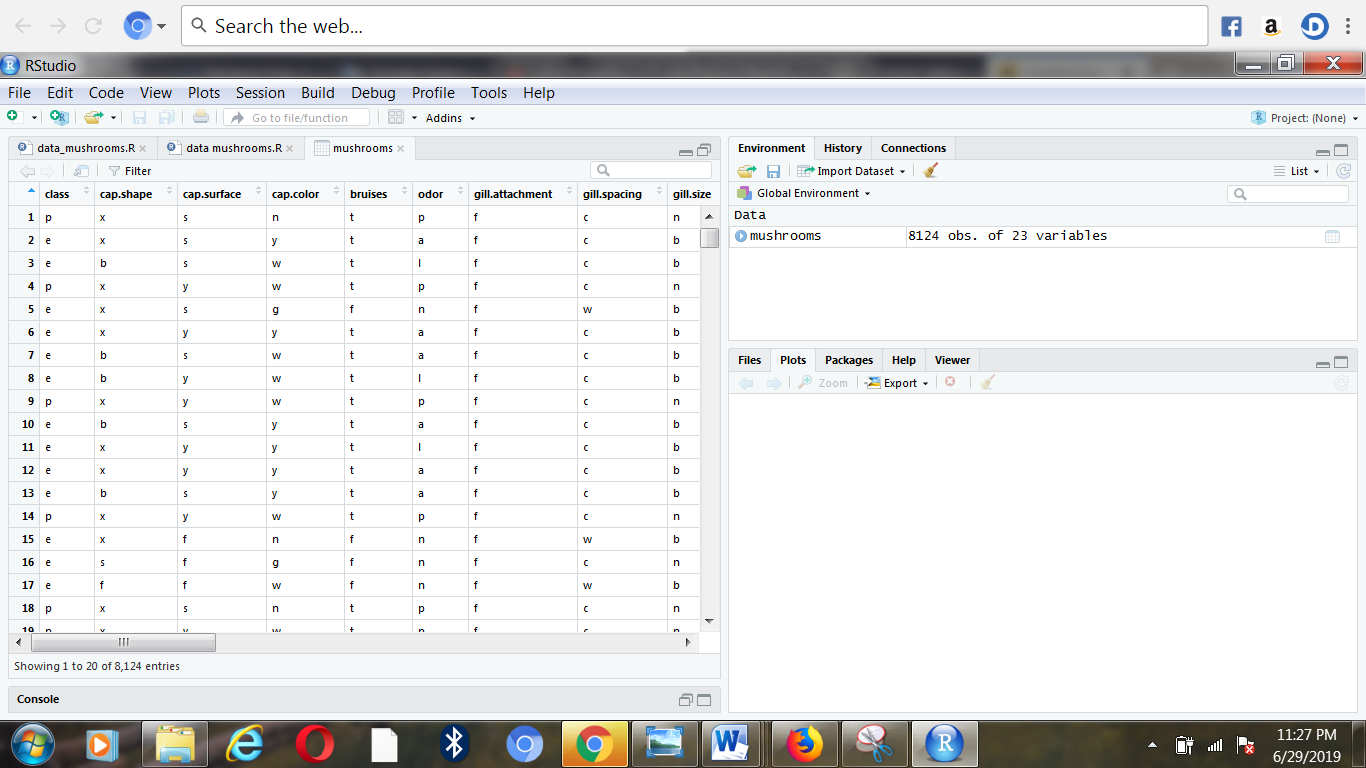
chi square

untuk mengetahui ada tidaknya hubungan atribut dengan label kelas.

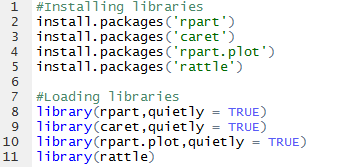
1. **Contoh Penerapan Analisis Decision Tree**

Pernyataan masalah : untuk mempelajari set data jamur untuk memprediksi apakah jamur yang diberikan dapat dimakan atau beracun bagi manusia.

Deskripsi set data : set data yang diberikan berisi total 8124 pengamatan dari berbagai jenis jamur dan sifat sifatnya seperti bau,habitat,populasi,dll.Struktur set data yang lebih mendalam ditampilkan dalam demo dibawah ini.



Tujuan analisis : Untuk membangun model Pohon Keputusan untuk mengklasifikasikan sampel jamur sebagai beracun atau dapat dimakan dengan mempelajari sifat-sifatnya seperti bau, akar, habitat, dll.  
langkah 1 : install dan load libraries



langkah 2 : Import data set

mushrooms <- read.csv("D:/SEMESTER 4 KULIAH/STATISTIK MULTIVARIAT/DATA DATA BIG DATA/cluster analysis kmeans/mushrooms.csv")

> View(mushrooms)

Sekarang, untuk menampilkan struktur set data, Anda dapat menggunakan fungsi R yang disebut str ():

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # structure of the data  str(mushrooms)    Output menunjukkan sejumlah variabel prediktor yang digunakan untuk memprediksi kelas output dari jamur(beracun atau dapat dimakan).  Langkah 3 : Pembersihan Data  Pada tahap ini, kita harus mencari nilai nol atau hilang dan variabel yang tidak perlu sehingga prediksi kita seakurat mungkin. Dalam cuplikan kode di bawah ini saya telah menghapus variabel ‘veil.type’ karena tidak berpengaruh pada hasil. Inkonsistensi dan data yang berlebihan tersebut harus diperbaiki pada langkah ini.  # number of rows with missing values  nrow(mushrooms) - sum(complete.cases(mushrooms))  [1] 0  # deleting redundant variable `veil.type`  mushrooms$veil.type <- NULL  Langkah 4: Eksplorasi dan Analisis Data Untuk mendapatkan pemahaman yang baik tentang 21 variabel prediktor, saya telah membuat tabel untuk setiap variabel prediktor vs tipe kelas (variabel respons / hasil) untuk memahami apakah variabel prediktor tertentu signifikan untuk mendeteksi output atau tidak.  Saya telah menunjukkan tabel hanya untuk variabel 'bau', Anda dapat melanjutkan dan membuat tabel untuk masing-masing variabel dengan mengikuti cuplikan kode di bawah ini:  # analyzing the odor variable  > table(mushrooms$class,mushrooms$odor)  a&nbsp; &nbsp; &nbsp;&nbsp;c&nbsp; &nbsp; &nbsp; f&nbsp; &nbsp; &nbsp; &nbsp;l&nbsp; &nbsp; &nbsp; &nbsp;m&nbsp; &nbsp; &nbsp; &nbsp;n&nbsp; &nbsp; &nbsp; &nbsp;p&nbsp; &nbsp; &nbsp; &nbsp; s&nbsp; &nbsp; &nbsp; y  e&nbsp; &nbsp;400&nbsp; &nbsp; &nbsp;0&nbsp; &nbsp; &nbsp; 0&nbsp; &nbsp; &nbsp;400&nbsp; &nbsp; &nbsp;0&nbsp; &nbsp; 3408&nbsp; &nbsp; 0&nbsp; &nbsp; &nbsp; &nbsp;0&nbsp; &nbsp; &nbsp; 0  p&nbsp; &nbsp;0&nbsp; &nbsp; &nbsp; 192&nbsp; &nbsp;2160&nbsp; 0&nbsp; &nbsp; &nbsp; 36&nbsp; &nbsp; &nbsp;120&nbsp; &nbsp; 256&nbsp; 576&nbsp; 576  # analyzing the odor variable  > table(mushrooms$class,mushrooms$odor)    a c f l m n p s y  e 400 0 0 400 0 3408 0 0 0  p 0 192 2160 0 36 120 256 576 576  Dalam cuplikan di atas, 'e' singkatan dari kelas yang dapat dimakan dan 'p' adalah singkatan dari kelas beracun dari jamur. Output di atas menunjukkan bahwa jamur dengan nilai bau ‘c’, ‘f’, ‘m’, ‘p’, s ’, s’, dan ‘y’ jelas beracun. Dan jamur yang memiliki almond (a) bau (400) dapat dimakan. Pengamatan seperti itu akan membantu kita untuk memprediksi kelas keluaran dengan lebih akurat. Langkah kami selanjutnya dalam tahap eksplorasi data adalah memprediksi variabel mana yang akan menjadi yang terbaik untuk memisahkan Pohon Keputusan. Karena alasan ini, saya telah merencanakan grafik yang mewakili pemisahan untuk masing-masing dari 21 variabel, hasilnya ditunjukkan di bawah ini:  number.perfect.splits <- apply(X=mushrooms[-1], MARGIN = 2, FUN = function(col){  + t <- table(mushrooms$class,col)  + sum(t == 0)  + })  >  > # Descending order of perfect splits  > order <- order(number.perfect.splits,decreasing = TRUE)  > number.perfect.splits <- number.perfect.splits[order]    > # Plot graph  > par(mar=c(10,2,2,2))  > barplot(number.perfect.splits,  + main="Number of perfect splits vs feature",  + xlab="",ylab="Feature",las=2,col="wheat")    Output menunjukkan bahwa variabel 'bau' memainkan peran penting dalam memprediksi kelas output jamur.  Langkah 5: Penyebaran Data Penyambungan Data adalah proses pemisahan data ke dalam satu set pelatihan dan satu set pengujian. Set pelatihan digunakan untuk membangun model Pohon Keputusan dan set uji digunakan untuk memvalidasi efisiensi model. Pemisahan dilakukan dalam cuplikan kode di bawah ini :   |  | | --- | | > #data splicing  > set.seed(12345)  > train <- sample(1:nrow(mushrooms),size = ceiling(0.80\*nrow(mushrooms)),replace = FALSE)  > # training set  > mushrooms\_train <- mushrooms[train,]  > # test set  > mushrooms\_test <- mushrooms[-train,] | |  | | |  | | --- | |  | |   Untuk membuat demo ini lebih menarik dan untuk meminimalkan jumlah jamur beracun yang salah diklasifikasikan sebagai dapat dimakan, kami akan memberikan penalti 10x lebih besar, daripada penalti untuk mengklasifikasikan jamur yang dapat dimakan sebagai beracun karena alasan yang jelas.  > # penalty matrix  > penalty.matrix <- matrix(c(0,1,10,0), byrow=TRUE, nrow=2)    Langkah 6: Membangun model Pada tahap ini, kami akan membangun Decision Tree dengan menggunakan algoritma rpart (Recursive Partitioning And Regression Trees):  > # building the classification tree with rpart  > tree <- rpart(class~.,  + data=mushrooms\_train,  + parms = list(loss = penalty.matrix),  + method = "class")    Langkah 7: Memvisualisasikan pohon  Pada langkah ini, kami akan menggunakan pustaka rpart.plot untuk merencanakan Pohon Keputusan terakhir kami:  > # Visualize the decision tree with rpart.plot  > rpart.plot(tree, nn=TRUE)  D:\SEMESTER 4 KULIAH\STATISTIK MULTIVARIAT\DATA DATA BIG DATA\cluster analysis kmeans\Rplot.png  Langkah 8: Menguji model  Sekarang untuk menguji model Decision Tree kami, kami akan menerapkan set data pengujian pada model kami seperti:  #Testing the model  > pred <- predict(object=tree,mushrooms\_test[-1],type="class")    Langkah 9: Menghitung akurasi  Kami akan menggunakan matriks kebingungan untuk menghitung keakuratan model. Ini kodenya:    Output menunjukkan bahwa semua sampel dalam set data uji telah diklasifikasikan dengan benar dan kami telah mencapai akurasi 100% pada set data uji dengan interval kepercayaan 95% (0,9977, 1). Dengan demikian kita dapat mengklasifikasikan jamur dengan benar baik beracun atau dapat dimakan menggunakan model Pohon Keputusan ini.  BAB III  PENUTUP  Kesimpulan  Penelitian menggunakan metode klasikasi dengan decision tree sangat efektif untuk digunakan bisa dilihat dari banyaknya kelebihan kelebihan sebaggai analisa suatu permasalahn suatu data.Namun dibalik kelebihan tersebut klasifikasi dengan metode decision tree juga mempunyai kelemahan  Saran  Diharapkan setelah membaca makalah ini para pembaca dapat mengetehui salah satu materi tentang analisis random forest dengan metode Decisin Tree (pohon keputusan) yang merupakan salah satu dari teknik statistika multivariat menggunakan R-Studio. Menyadari bahwa makalah belum dari kata sempurna, kedepannya kami akan lebih fokus dan detail dalam menjelaskan tentang makalah di atas dengan sumber-sumber yang lebih banyak, yang tentunya dapat di pertanggungjawabkan. |
|  |
| |  | | --- | |  | |

# **DAFTAR PUSTAKA**

[**https://datascience.or.id/article/Perbedaan-Supervised-and-Unsupervised-Learning-5a8fa6e6**](https://datascience.or.id/article/Perbedaan-Supervised-and-Unsupervised-Learning-5a8fa6e6)

[**https://www.academia.edu/7139996/Supervised\_Learning**](https://www.academia.edu/7139996/Supervised_Learning)

[**https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d**](https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d)