**Pokok Bahasan III**

**Decision tree**

**Kode Pokok Bahasan**: TIK.RPL03.001.003.01

**Deskripsi Pokok Bahasan**:

Membahas bagaimana pembuatan Decision Tree pada dataset yang diberikan.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Elemen Kompetensi | Indikator Kinerja | Jml Jam | Hal |
| 1 | Menerapkan Decision Tree pada dataset Car Evaluation. | 1.1 Membuat Decision Tree menggunakan Library Rpart dan Rpart Plot menggunakan dataset Car Evaluation | 1 | 12 |
| 2 | Menerapkan prediksi dan class pada Decision Tree data Weather Nominal dataset 2 dan dataset Weather Nominal 1. | 1.1 Membuat Decision Tree menggunakan Library Rpart dan Rpart Plot menggunakan data Weather Nominal  1.2 Membuat prediksi dari Decision Tree  1.3 Membuat class dari prediksi | 2 | 15 |

**TUGAS PENDAHULUAN**

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal R pada PC masing-masing praktikan.

2. Menginstal R Studio pada PC masing-masing praktikan.

**DAFTAR PERTANYAAN**

1. Apa itu decision tree?
2. Apa kegunaan Decision tree?
3. Berikan salah satu contoh dari decision tree suatu data!

1. Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko.

2. 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.

3. Decision tree digunakan untuk menangani kumpulan data non-linier secara efektif. Decision tree digunakan dalam kehidupan nyata di banyak bidang, seperti teknik, perencanaan sipil, hukum, dan bisnis. Decision tree dapat dibagi menjadi dua jenis; variabel kategori dan pohon keputusan variabel kontinu.

**TEORI SINGKAT**

RStudio adalah lingkungan pengembangan terintegrasi sumber terbuka dan gratis untuk R, bahasa pemrograman untuk komputasi statistik dan grafik. RStudio didirikan oleh J.J.Allaire, pencipta bahasa pemrograman ColdFusion. Hadley Wickham adalah Kepala Ilmuwan di RStudio.

Data preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur.

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya.

**LAB SETUP**

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.

2. Menjalankan R Studio.

**ELEMEN KOMPETENSI I**

**Deskripsi:**

Menerapkan Decision Tree pada data Car Evaluation.

**Kompetensi Dasar**:

Membuat Decision Tree menggunakan Library Rpart dan Rpart Plot menggunakan data Car Evaluation.

**Latihan 1.1.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta untuk membangun decision tree menggunakan library yang disediakan oleh R.

**Langkah-Langkah Praktikum:**

**1.** Disediakan data sebagai berikut :

**Link Dataset:**

**https://www.kaggle.com/code/prashant111/decision-tree-classifier-tutorial/data**

**2.** Input data ke R studio

> library(rpart)

> dataku\_namapraktikan <-read.csv("C:/Users/Sectio/Downloads/car\_evaluation.csv")

> View(dataku\_namapraktikan)

Table

Description automatically generated

3.Lakukan Preprocessing untuk mengganti nama kolom pada dataset

#preprocessing dataset

> colnames(thedata) <- c('buying', 'maint', 'doors', 'persons', 'lug\_boot', 'safety', 'class')

Table

Description automatically generated

3. Buatlah decision tree dari data input.

> pohon <- rpart(class~.,data = dataku\_namapraktikan, method = "class", control=rpart.control(minsplit = 2, cp= 0))

> library(rpart.plot)

> prp(pohon,extra = 1)

Output :

|  |
| --- |
|  |

4. Berikan penjelasan tentang output yang muncul!

**Interpretasi output decision tree :**

|  |
| --- |
| Dari hasil yang kita dapatkan karna data yang mungkin terlalu banyak jadi tidak terlalu bisa dilihat data yang ditampilkan, tetapi penjelasan yang umum bisa diambil dari gambar tersebut adalah bagaimana ketika kita mempunyai suatu data dan kemudian kita membutuhkan mapping yang lebih mudah untuk dibaca ketika dibutuhkan, maka decision tree ini lah salah satu opsinya, untuk menentukan pilihan-pilihan yang kita inginkan. |

**ELEMEN KOMPETENSI II**

**Deskripsi:**

Menerapkan prediksi menggunakan Decision Tree pada data Weather Nominal

dataset 2.

**Kompetensi Dasar**:

Membuat prediksi pada Tree menggunakan data Weather Nominal dataset 2.

**Latihan 1.2.1**

**Penjelasan Singkat :**

Pada latihan ini anda akan diminta untuk membangun decision tree dan melakukan prediksi menggunakan library yang disediakan oleh R.

**Langkah-Langkah Praktikum:**

**1.** Disediakan data sebagai berikut :

Dataset 1 :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **cuaca** | **suhu** | **kelembaban** | **berangin** | **bermain** |
| cerah | panas | Tinggi | salah | tidak |
| cerah | panas | Tinggi | benar | tidak |
| berawan | panas | Tinggi | salah | ya |
| hujan | sejuk | Tinggi | salah | ya |
| hujan | dingin | Normal | salah | ya |
| hujan | dingin | Normal | benar | ya |
| berawan | dingin | Normal | benar | ya |
| cerah | sejuk | Tinggi | salah | tidak |
| cerah | dingin | Normal | salah | ya |
| hujan | sejuk | Normal | salah | ya |
| cerah | sejuk | Normal | benar | ya |
| berawan | sejuk | Tinggi | benar | ya |
| berawan | panas | Normal | salah | ya |
| hujan | sejuk | Tinggi | benar | tidak |

**Data set 2 :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cuaca | suhu | kelembaban | angin | bermain |
| cerah | 69 | 70 | biasa | ya |
| cerah | 72 | 95 | biasa | tidak |
| cerah | 75 | 70 | kencang | ya |
| cerah | 80 | 90 | kencang | tidak |
| cerah | 85 | 85 | biasa | tidak |
| hujan | 65 | 70 | kencang | tidak |
| hujan | 68 | 80 | biasa | ya |
| hujan | 70 | 96 | biasa | ya |
| hujan | 71 | 80 | kencang | tidak |
| hujan | 75 | 80 | biasa | ya |
| mendung | 64 | 65 | kencang | ya |
| mendung | 72 | 90 | kencang | ya |
| mendung | 81 | 75 | biasa | ya |
| mendung | 83 | 78 | biasa | ya |

**2.** Input data ke R studio

> dataku2\_namapraktikan = read.delim(‘clipboard’)

> library(tree)

> dataku2\_namapraktikan$bermain = as.factor(dataku2\_namapraktikan $bermain)

**Latihan 1.2.2**

3. Buatlah decision tree dari data input.

sample=read.csv('dataset1.csv', header=TRUE, sep=";")

samplegolf<-sample(1:nrow(sample),0.80\*nrow(sample))

traininggolf<-data.frame(sample)[samplegolf,]

testinggolf<-data.frame(sample)[-samplegolf,]

pohongolf <- rpart(bermain ~ cuaca + suhu + kelembaban + berangin,data=traininggolf, method = "class", control = rpart.control(minsplit = 5, cp = 0))

prp(pohongolf,extra=4,box.col=c("pink","red","blue"))

Output :

|  |
| --- |
| Output yang bisa didapat dari tree diatas adalah seberapa banyak warga yang bermain golf tergantung dari cuaca yang sedang terjadi, ada lebih banyak yang memilih untuk bermain ketika cuaca lembab dan sejuk disbanding cerah dan panas. |

**Latihan 1.2.3**

4. Buat prediksi dengan fungsi predict

prediksi=predict(pohongolf,testinggolf)

pred.respon<- colnames(prediksi)[max.col(prediksi, ties.method = c("random"))]

pred.respon

5. Buat class table.

class=table(pred.respon,testinggolf$bermain)

class

6. Berikan penjelasan tentang output yang muncul!

**Interpretasi output decision tree :**

|  |
| --- |
|  |

**CEK LIST**

|  |  |  |  |
| --- | --- | --- | --- |
| Elemen Kompetensi | No Latihan | Penyelesaian | |
| Selesai | Tidak selesai |
| 1 | 1.1.1 |  |  |
| 2 | 1.2.1 |  |  |
| 1.2.2 |  |  |
| 1.2.3 |  |  |

**FORM UMPAN BALIK**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Elemen Kompetensi** | **Tingkat Kesulitan** | | | **Tingkat Ketertarikan** | | | **Waktu Penyelesaian dalam menit** |
| Memahami data pre-processing. |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |
|  |  |  |  |  |  |  |
| Mengimplementasikan pre-processing data. |  |  |  |  |  |  |  |
|  |  | Sangat Mudah |  |  | Tidak Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Mudah |  |  | Cukup Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | * Biasa |  |  | Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sulit |  |  | * Sangat Tertarik |  |
|  |  |  |  |  |  |  |
|  |  | Sangat Sulit |  |  |  |  |
|  |  |  |  |  |  |  |