**Praktikum 3**

**Decision Tree Menggunakan Data Numerik**

**Data weather Numerik**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

1. **Menghitung entropy dan gain dari data weather nominal**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Node1 | jumlah kasus | ya | tidak | entropy | gain |
| Total | 14 | 9 | 5 | 0.94028596 |  |
|  |  |  |  |  |  |
| suhu |  |  |  |  | 0.045334172 |
| <= 70 | 5 | 4 | 1 | 0.72192809 |  |
| > 70 | 9 | 5 | 4 | 0.99107606 |  |
|  |  |  |  |  |  |
| suhu |  |  |  |  | 0.025078174 |
| <= 75 | 10 | 7 | 3 | 0.8812909 |  |
| > 75 | 4 | 2 | 2 | 1 |  |
|  |  |  |  |  |  |
| suhu |  |  |  |  | 0.000489469 |
| <= 80 | 11 | 7 | 4 | 0.9456603 |  |
| > 80 | 3 | 2 | 1 | 0.91829583 |  |
|  |  |  |  |  |  |
| kelembaban |  |  |  |  | 0.01495607 |
| <= 70 | 4 | 3 | 1 | 0.81127812 |  |
| > 70 | 10 | 6 | 4 | 0.97095059 |  |
|  |  |  |  |  |  |
| kelembaban |  |  |  |  | 0.045334172 |
| <= 75 | 5 | 4 | 1 | 0.72192809 |  |
| > 75 | 9 | 5 | 4 | 0.99107606 |  |
|  |  |  |  |  |  |
| kelembaban |  |  |  |  | 0.102243564 |
| <= 80 | 9 | 7 | 2 | 0.76420451 |  |
| > 80 | 5 | 2 | 3 | 0.97095059 |  |
|  |  |  |  |  |  |
| kelembaban |  |  |  |  | 0.025078174 |
| <= 85 | 10 | 7 | 3 | 0.8812909 |  |
| > 85 | 4 | 2 | 2 | 1 |  |
|  |  |  |  |  |  |
| Cuaca |  |  |  |  | 0.593517889 |
| cerah | 5 | 2 | 3 | 0.97095059 |  |
| mendung | 4 | 4 | 0 |  |  |
| hujan | 5 | 3 | 2 | 0.97095059 |  |
|  |  |  |  |  |  |
| Angin |  |  |  |  | 0.04812703 |
| biasa | 8 | 6 | 2 | 0.81127812 |  |
| kencang | 6 | 3 | 3 | 1 |  |

1. **Membuat Decision Tree menggunakan Library Rpart dan Rpart Plot menggunakan data Weather Nominal**

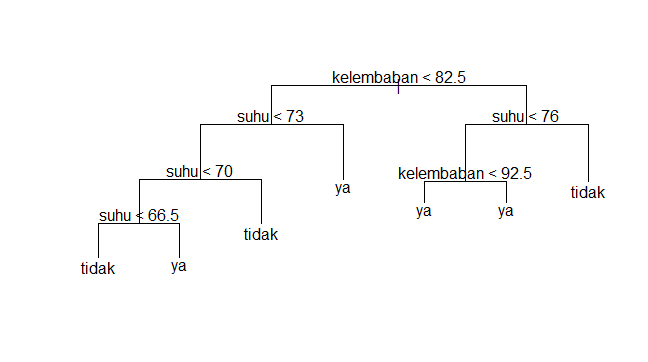
> library(tree)

> tes$bermain = as.factor(tes$bermain)

> tree2 = tree(bermain~kelembaban + suhu, data = tes, split = c("gini"), control = tree.control(nobs = 26, mincut = 1, minsize = 2))

> plot(tree2)

> text(tree2)



**Keterangan:**

|  |
| --- |
| Pembagi utama kelembaban udara kurang dari 82.5 terbagi menjadi dua yaitu suhu dibawah 73 dan suhu di bawah 76  Pada factor suhu bukan kurang dari 73 maka bermain,jika dibawah 73 maka dipengaruhi suhu kurang dari 70.  Pada factor suhu kurang dari 70,jika tidak maka tidak bermain. Jika ya maka dipengaruhi suhu kurang dari 66.5  Pada suhu kurang dari 66.5 maka tidak bermain, selain itu bermain.  Jika suhu bukan dibawah 76maka tidak bermain, jika dibawah 76 maka dipengaruhi kelembaban kurang dari 92.5  Baik kelembaban dibawah 92,5 maupun diatas 92.5 keduanya akan bermain. |

**TUGAS**

Buatlah decision tree menggunakan library tree seperti di atas menggunakan data datapraktikum3.

(DATA BARU dengan Jenis Beras 18/9/2018)

> dataku=read.delim("clipboard")

> library(rpart)

> library(rpart.plot)

> library(caret)

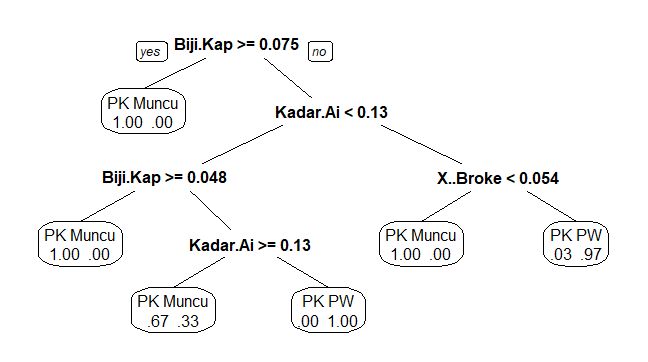
> sampel<-sample(1:nrow(dataku),0.75\*nrow(dataku))

> training<-data.frame(dataku)[sampel,]

> testing<-data.frame(dataku)[-sampel,]

> pohon=rpart(Jenis.Beras~., data=training,method="class",control = rpart.control(minsplit=5,cp=0))

> prp(pohon,extra=4)



> prediksi=predict(pohon,testing)

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

> install.packages("e1071")

> library(e1071)

> class=table(pred.respon,testing$Jenis.Beras)

> confusionMatrix(class)

Confusion Matrix and Statistics

pred.respon PK Muncul PK PW

PK Muncul 4 3

PK PW 1 8

Accuracy : 0.75

95% CI : (0.4762, 0.9273)

No Information Rate : 0.6875

P-Value [Acc > NIR] : 0.4069

Kappa : 0.4754

Mcnemar's Test P-Value : 0.6171

Sensitivity : 0.8000

Specificity : 0.7273

Pos Pred Value : 0.5714

Neg Pred Value : 0.8889

Prevalence : 0.3125

Detection Rate : 0.2500

Detection Prevalence : 0.4375

Balanced Accuracy : 0.7636

'Positive' Class : PK Muncul