# Penggunaan Class *myID3* sebagai Classifier

## Praktikum 1 IF4071

## Semester I Tahun 2014 / 2015

## Kelompok WbTeladan


oleh :

Iskandar Setiadi / 13511073

Raden Fajar Hadria Putra / 13511076

M Andri Eka Fauzy / 13511088

Sekolah Teknik Elektro dan Informatika (STEI ITB)

Institut Teknologi Bandung

Jl. Ganesha No. 10, Bandung 40132

Tahun 2014

## A. Dataset

Dataset yang Digunakan:

Dataset 1 (Dataset Awal):

```
@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

Dataset 2 (Dataset Awal + 1 Data Noise):

```
@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
sunny,hot,normal,TRUE,no
```

Dataset 3 (Dataset Komplemen dari Dataset 2):

```
@relation weather.symbolic
```

```
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,normal,FALSE,?
sunny,mild,high,TRUE,?
sunny,mild,normal,FALSE,?
sunny,cool,high,TRUE,?
sunny,cool,high,FALSE,?
sunny,cool,normal,TRUE,?
overcast,hot,high,TRUE,?
overcast,hot,normal,TRUE,?
overcast,mild,high,FALSE,?
overcast,mild,normal,TRUE,?
overcast,mild,normal,FALSE,?
overcast,cool,high,TRUE,?
overcast,cool,high,FALSE,?
overcast,cool,normal,FALSE,?
rainy,hot,high,TRUE,?
rainy,hot,high,FALSE,?
rainy,hot,normal,TRUE,?
rainy,hot,normal,FALSE,?
rainy,mild,high,TRUE,?
rainy,mild,normal,TRUE,?
rainy,cool,high,FALSE,?
```

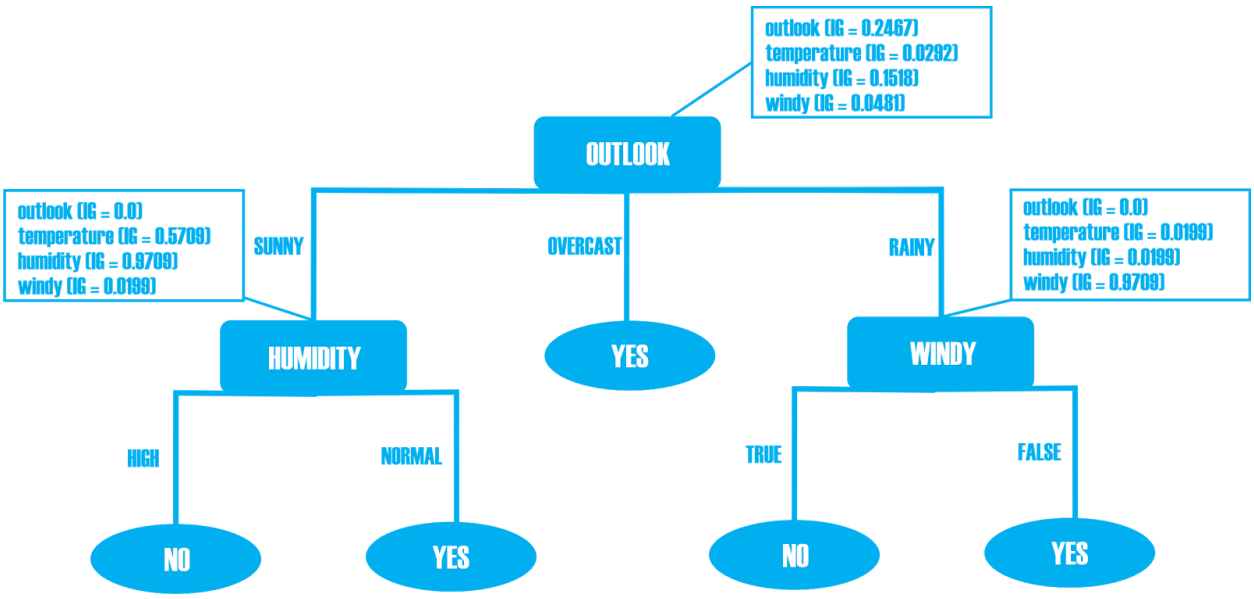
## B. Hasil Pembelajaran

Hasil Pembelajaran dengan Model dari Dataset 1 :

```
|
|
 outlook (IG = 0.0)
 temperature (IG = 0.5709505944546686)
 humidity (IG = 0.9709505944546686)
 windy (IG = 0.01997309402197489)
 humidity = high
 Kelas : yes [LEAF
-- ID3 Model --
|
 outlook (IG = 0.2467498197744391)
 temperature (IG = 0.029222565658954647)
 humidity (IG = 0.15183550136234136)
 windy (IG = 0.04812703040826927)
 outlook = sunny]

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.5709505944546686)
 humidity (IG = 0.9709505944546686)
 windy (IG = 0.01997309402197489)
 humidity = normal
 Kelas : no [LEAF]



|
 outlook (IG = 0.2467498197744391)
 temperature (IG = 0.029222565658954647)
 humidity (IG = 0.15183550136234136)
```

```
 windy (IG = 0.04812703040826927)
 outlook = overcast
 Kelas : no [LEAF]


|
 outlook (IG = 0.2467498197744391)
 temperature (IG = 0.029222565658954647)
 humidity (IG = 0.15183550136234136)
 windy (IG = 0.04812703040826927)
 outlook = rainy

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.01997309402197489)
 humidity (IG = 0.01997309402197489)
 windy (IG = 0.9709505944546686)
 windy = TRUE
 Kelas : yes [LEAF]

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.01997309402197489)
 humidity (IG = 0.01997309402197489)
 windy (IG = 0.9709505944546686)
 windy = FALSE
 Kelas : no [LEAF]
```

Ilustrasi Pohon Model 1 :

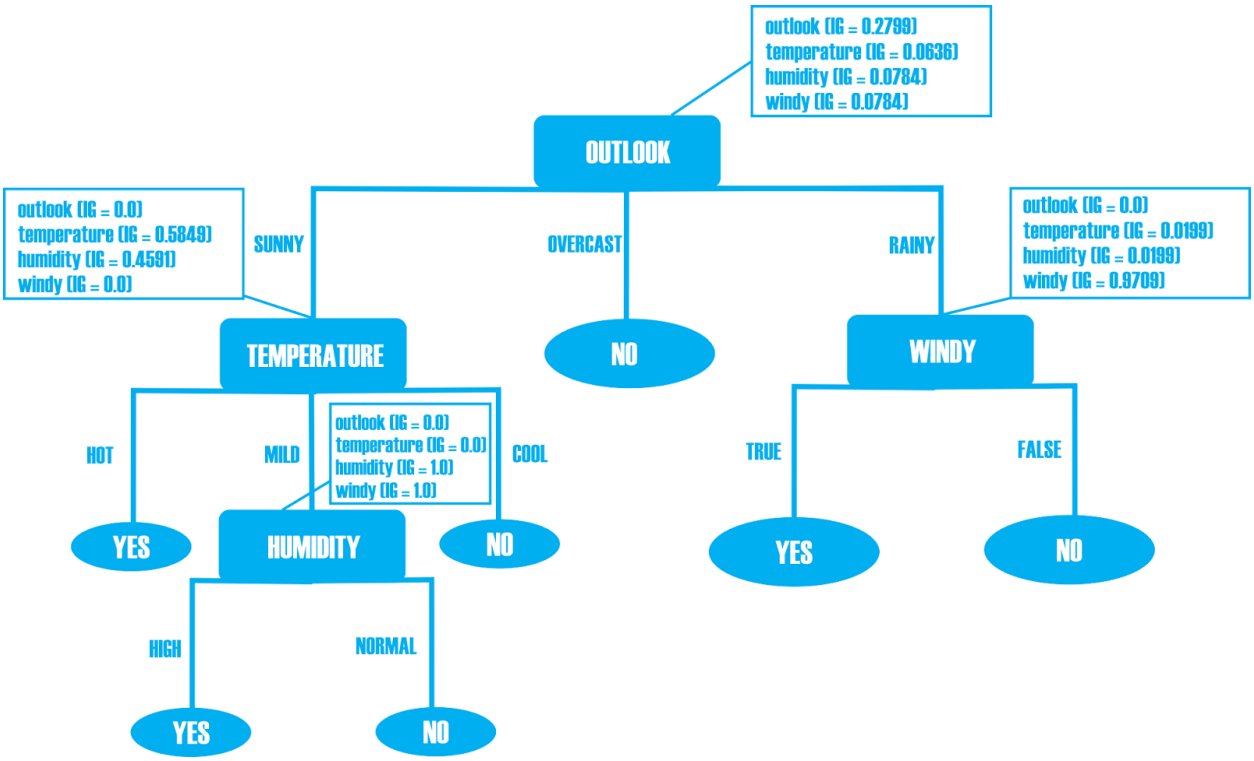



Hasil Pembelajaran dengan Model dari Dataset 2 :

```
-- ID3 Model --
|
 outlook (IG = 0.27998206268131665)
 temperature (IG = 0.06364122949221451)
 humidity (IG = 0.078495797927147)
 windy (IG = 0.078495797927147)
 outlook = sunny

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.5849625007211563)
 humidity (IG = 0.4591479170272448)
 windy (IG = 0.0)
 temperature = hot
 Kelas : yes [LEAF]

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.5849625007211563)
 humidity (IG = 0.4591479170272448)
 windy (IG = 0.0)
 temperature = mild

|
|
|
 outlook (IG = 0.0)
 temperature (IG = 0.0)
 humidity (IG = 1.0)
 windy (IG = 1.0)
 humidity = high
 Kelas : yes [LEAF]

|
```

```
|
|
 outlook (IG = 0.0)
 temperature (IG = 0.0)
 humidity (IG = 1.0)
 windy (IG = 1.0)
 humidity = normal
 Kelas : no [LEAF]



|
|
 outlook (IG = 0.0)
 temperature (IG = 0.5849625007211563)
 humidity (IG = 0.4591479170272448)
 windy (IG = 0.0)
 temperature = cool
 Kelas : no [LEAF]



|
 outlook (IG = 0.27998206268131665)
 temperature (IG = 0.06364122949221451)
 humidity (IG = 0.078495797927147)
 windy (IG = 0.078495797927147)
 outlook = overcast
 Kelas : no [LEAF]

|
 outlook (IG = 0.27998206268131665)
 temperature (IG = 0.06364122949221451)
 humidity (IG = 0.078495797927147)
 windy (IG = 0.078495797927147)
 outlook = rainy

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.01997309402197489)
 humidity (IG = 0.01997309402197489)
 windy (IG = 0.9709505944546686)
 windy = TRUE
 Kelas : yes [LEAF]

|
|
 outlook (IG = 0.0)
 temperature (IG = 0.01997309402197489)
 humidity (IG = 0.01997309402197489)
 windy (IG = 0.9709505944546686)
 windy = FALSE
 Kelas : no [LEAF]
```

Ilustrasi Pohon Model 2 :




## C. Hasil Klasifikasi

Klasifikasi menggunakan dataset 3 sebagai test-set standar

Hasil klasifikasi dengan model dari dataset 1 :

myID3

```
Results
======

Correctly Classified Instances          21               100      %
Incorrectly Classified Instances    0               0       %
Kappa statistic                     1
Mean absolute error                 0
Root mean squared error             0
Relative absolute error             0       %
Root relative squared error         0       %
Total Number of Instances           21
```

ID3 WEKA

```
Results
======

Correctly Classified Instances          21               100      %
Incorrectly Classified Instances    0               0       %
Kappa statistic                     1
```

```
Mean absolute error                     0
Root mean squared error                 0
Relative absolute error                 0       %
Root relative squared error             0       %
Total Number of Instances                       21
```

Pengisian Atribut Play Dataset 3 oleh Model 1 :

```
sunny,hot,normal,FALSE,?
Classifying result:
outlook = sunny
humidity = normal
no

sunny,mild,high,TRUE,?
Classifying result:
outlook = sunny
humidity = high
yes

sunny,mild,normal,FALSE,?
Classifying result:
outlook = sunny
humidity = normal
no

sunny,cool,high,TRUE,?
Classifying result:
outlook = sunny
humidity = high
yes

sunny,cool,high,FALSE,?
Classifying result:
outlook = sunny
humidity = high
yes

sunny,cool,normal,TRUE,?
Classifying result:
outlook = sunny
humidity = normal
no

overcast,hot,high,TRUE,?
Classifying result:
outlook = overcast
no

overcast,hot,normal,TRUE,?
Classifying result:
outlook = overcast
no

overcast,mild,high,FALSE,?
Classifying result:
outlook = overcast
no
```

```
overcast,mild,normal,TRUE,?
Classifying result:
outlook = overcast
no

overcast,mild,normal,FALSE,?
Classifying result:
outlook = overcast
no

overcast,cool,high,TRUE,?
Classifying result:
outlook = overcast
no

overcast,cool,high,FALSE,?
Classifying result:
outlook = overcast
no

overcast,cool,normal,FALSE,?
Classifying result:
outlook = overcast
no

rainy,hot,high,TRUE,?
Classifying result:
outlook = rainy
windy = TRUE
yes

rainy,hot,high,FALSE,?
Classifying result:
outlook = rainy
windy = FALSE
no

rainy,hot,normal,TRUE,?
Classifying result:
outlook = rainy
windy = TRUE
yes

rainy,hot,normal,FALSE,?
Classifying result:
outlook = rainy
windy = FALSE
no

rainy,mild,high,TRUE,?
Classifying result:
outlook = rainy
windy = TRUE
yes

rainy,mild,normal,TRUE,?
Classifying result:
outlook = rainy
windy = TRUE
yes

rainy,cool,high,FALSE,?
Classifying result:
outlook = rainy
```

```
windy = FALSE
no
```

Hasil klasifikasi dengan model dari dataset 2 :

myID3

```
Results
======

Correctly Classified Instances          18               85.7143 %
Incorrectly Classified Instances     3               14.2857 %
Kappa statistic                          0.6667
Mean absolute error                      0.1429
Root mean squared error                  0.378
Relative absolute error                 26.9841 %
Root relative squared error             70.5279 %
Total Number of Instances               21
```

ID3 WEKA

```
Results
======

Correctly Classified Instances          18               85.7143 %
Incorrectly Classified Instances     3               14.2857 %
Kappa statistic                          0.6667
Mean absolute error                      0.1429
Root mean squared error                  0.378
Relative absolute error                 30.3571 %
Root relative squared error             79.0912 %
Total Number of Instances               21
```

## D. Kesimpulan

Dari hasil klasifikasi yang didapatkan oleh kedua model, dapat disimpulkan bahwa model-2 memiliki tingkat akurasi lebih kecil. Hal ini dikarenakan adanya *noise* pada dataset yang dijadikan *training set* yang membuat penghitungan *information gain* berbeda dari dataset 1 yang diasumsikan dan terbukti menghasilkan model dengan tingkat akurasi 100%. Dari *data noise* <sunny, hot, normal, TRUE, no>, model klasifikasi yang terbentuk pada *level 1* (**outlook** = sunny) mengalami perubahan akibat perbedaan nilai *information gain* tersebut.

Pada model 1, **temperature** memiliki IG sebesar 0.5709 sedangkan **humidity** memiliki IG sebesar 0.9709. Pada model 2, **temperature** memiliki IG sebesar 0.5849 sedangkan

**humidity** memiliki IG sebesar 0.4591. Hal ini menyebabkan perubahan hasil klasifikasi pada 3 *instances*, yaitu:

- <sunny, hot, normal, FALSE, ?>: model 1 = no, model 2 = yes
- <sunny, cool, high, TRUE, ?>: model 1 = yes, model 2 = no
- <sunny, cool, high, FALSE, ?>: model 1 = yes, model 2 = no

Eksperimen ini menunjukkan bahwa pembelajaran menggunakan *decision tree* dapat menangani data *noise*, dengan *tradeoff* berupa tingkat akurasi yang biasanya lebih buruk jika dibandingkan dengan model tanpa data *noise*.