

# Voter Classification Using C4.5

Richard Beindorff

Math 3220

# Problem

- Classify Voters based on Biographical Data

President	Age	Age Category	Years of Education	Number of Degrees	Gender
2	44	2.00	98	0	2
3	25	1.00	98	1	2
3	40	2.00	20	4	2
1	47	3.00	20	4	2

(Continued)

# Analysis

Age                      Age Category

- $\leq 34$         = Category 1
- $35 \leq 44$    = Category 2
- $45 \leq 64$    = Category 3
- $65+$         =category 4.

President	Age	Age Category	Years of Education	Number of Degrees	Gender
2	44	2.00	98	0	2
3	25	1.00	98	1	2
3	40	2.00	20	4	2
1	47	3.00	20	4	2

# Analysis

## Correlation coefficient

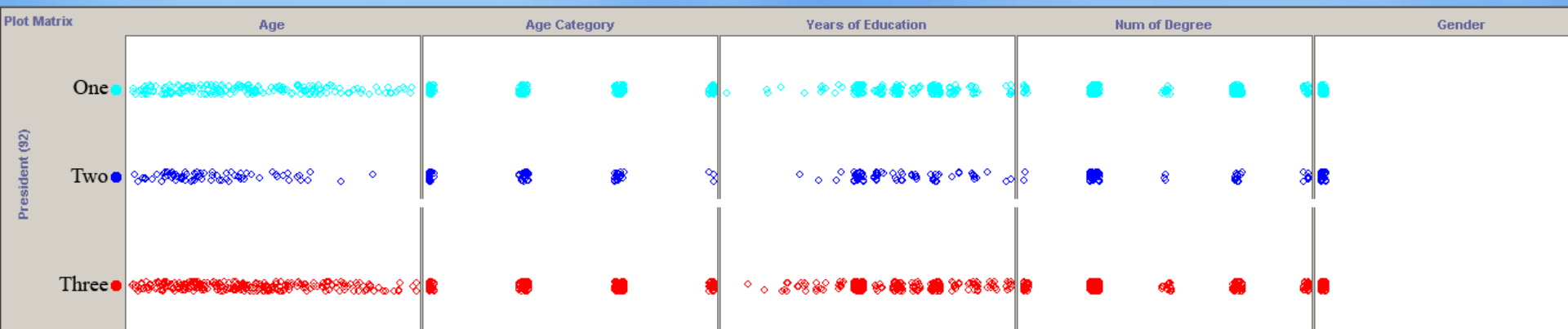
- Interdependence of two variables
- 1 = Perfect Correlation
- -1 = Perfect Negative

	Pres (92)	Age	Age Category	Years of Education	Num of Degrees	Gender
Pres (92)	1.0000	0.0297	0.0447	-0.0082	-0.0108	0.1024
Age	0.0297	1.0000	0.9393	-0.2143	-0.2144	0.0126
Age Category	0.0447	0.9393	1.0000	-0.1848	-0.1713	0.0009
Years of Education	-0.0082	-0.2143	-0.1848	1.0000	0.6003	-0.0122
Num of Degrees	-0.0108	-0.2144	-0.1713	0.6003	1.0000	-0.0269
Gender	0.1024	0.0126	0.0009	-0.0122	-0.0269	1.0000

# Analysis

Visualize

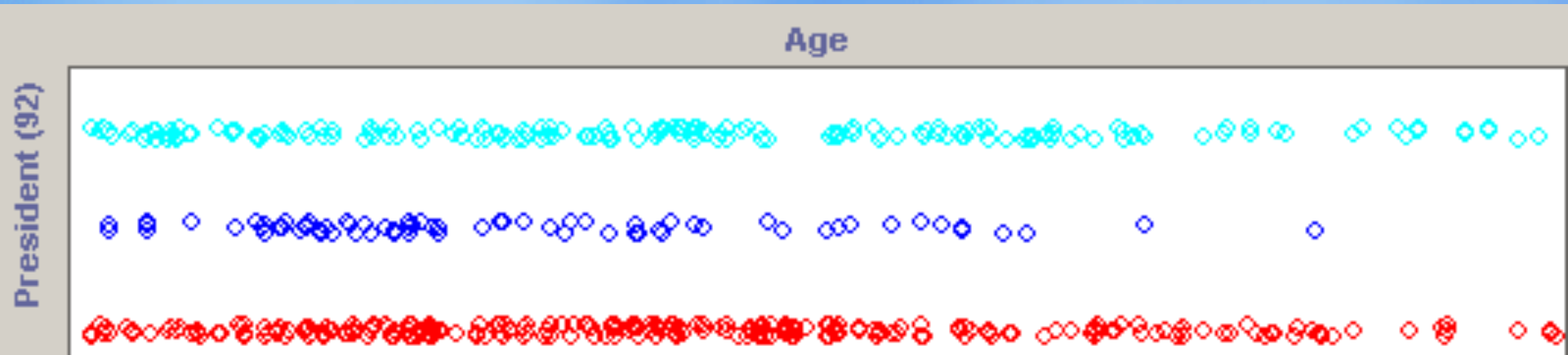
- Weka



# Analysis

Visualize

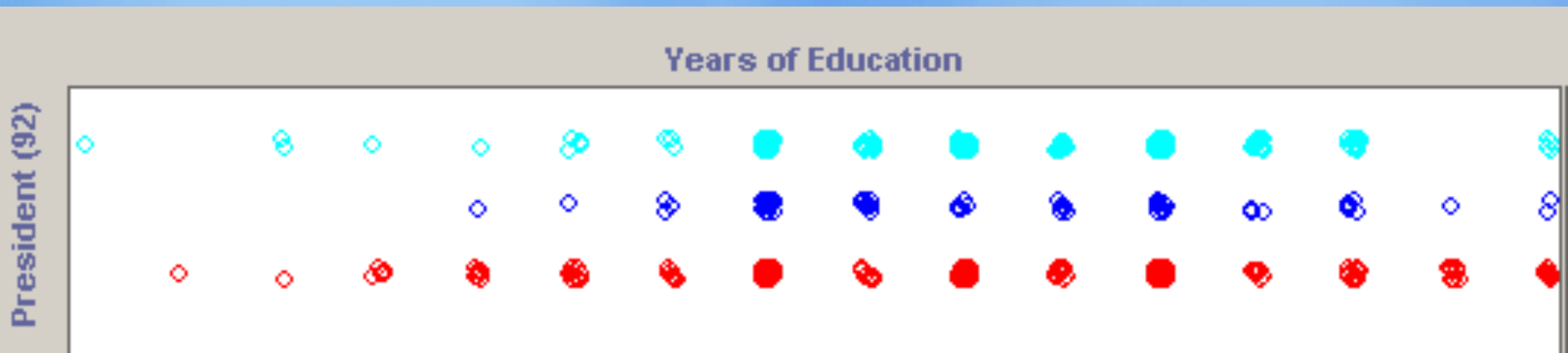
- Weka



# Analysis

Visualize

- Weka



- Clustering



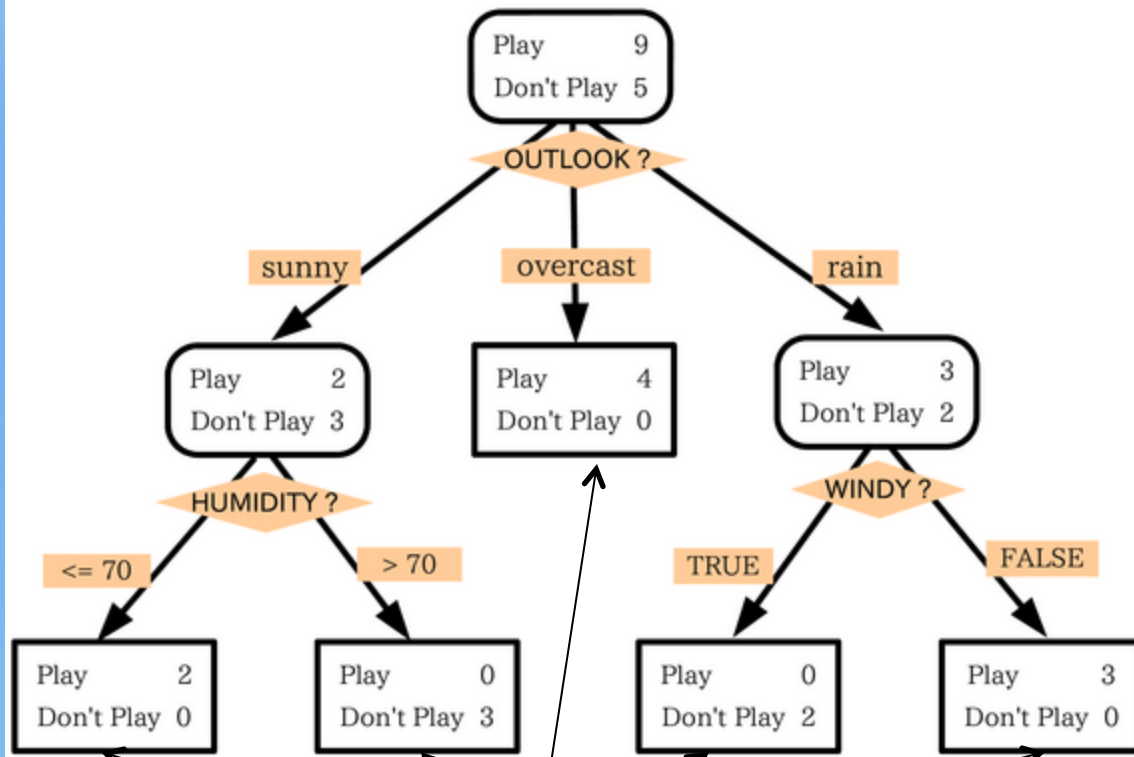
# Analysis

## C4.5 – Decision Tree

Size = Number of nodes

(8)

Dependent variable: PLAY



(\*)

Leaves (5)



# Analysis

C4.5

Accuracy

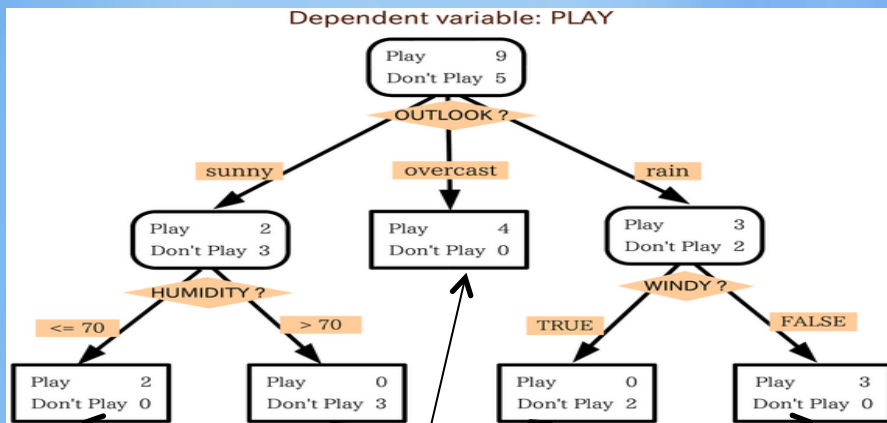
- All Attributes = 54.7374 %
- Removed: “Gender” = 54.4126 %
- Removed: “Age” = 51.7596 %
- Removed: “Age Category” = 56.1992 %
- Removed: “Years of Education” = 52.2469 %
- Removed: “Num of Degree” = **56.5241 %**

# Analysis

C4.5

Accuracy

- Removed: “Num of Degree” = **56.5241 %**
- Tree size = 67
- Leaves = 34



Size = Number of nodes

(8)

(\*)

Leaves

(5)

# Analysis

## C4.5 - Changed Inputs

- “Confidence Factor” increased from 0.25 to 1.0
- Confidence factor used for pruning tree.
- Small values mean more pruning (Accuracy).
- Ran again with accuracy = 57.9 %.
- Size = 139
- Leaves = 70

# Results

- **Best result:**
- “Num of Degree” = **56.5241 %**
- Tree size = 67
- Leaves = 34
- **Inconclusive**
- **Never exceeded 60%**
- **Recommend additional information be collected for better accuracy**



# References

- (\*)[http://upload.wikimedia.org/wikipedia/commons/f/ff/Decision\\_tree\\_model.png](http://upload.wikimedia.org/wikipedia/commons/f/ff/Decision_tree_model.png)