**WEKA Classification Analysis**

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: AnotherEmpRetentionData

Instances: 200

Attributes: 5

Age

Gender

Marital

HowApplied

Stayed2?

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

------------------

HowApplied = OnlineAd

| Age <= 40: N (35.0/5.0)

| Age > 40: Y (24.0/8.0)

HowApplied = PaperAd: Y (26.0/4.0)

HowApplied = WOM: Y (62.0/2.0)

HowApplied = Recruiter: Y (32.0/2.0)

HowApplied = JobFair: N (21.0/3.0)

Number of Leaves : 6

Size of the tree : 8

Time taken to build model: 0 seconds.

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 169 84.5 %

Incorrectly Classified Instances 31 15.5 %

Kappa statistic 0.6301

Mean absolute error 0.2105

Root mean squared error 0.3384

Relative absolute error 48.2767 %

Root relative squared error 72.5278 %

Total Number of Instances 200

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.688 0.081 0.800 0.688 0.739 0.634 0.870 0.747 N

0.919 0.313 0.862 0.919 0.890 0.634 0.870 0.898 Y

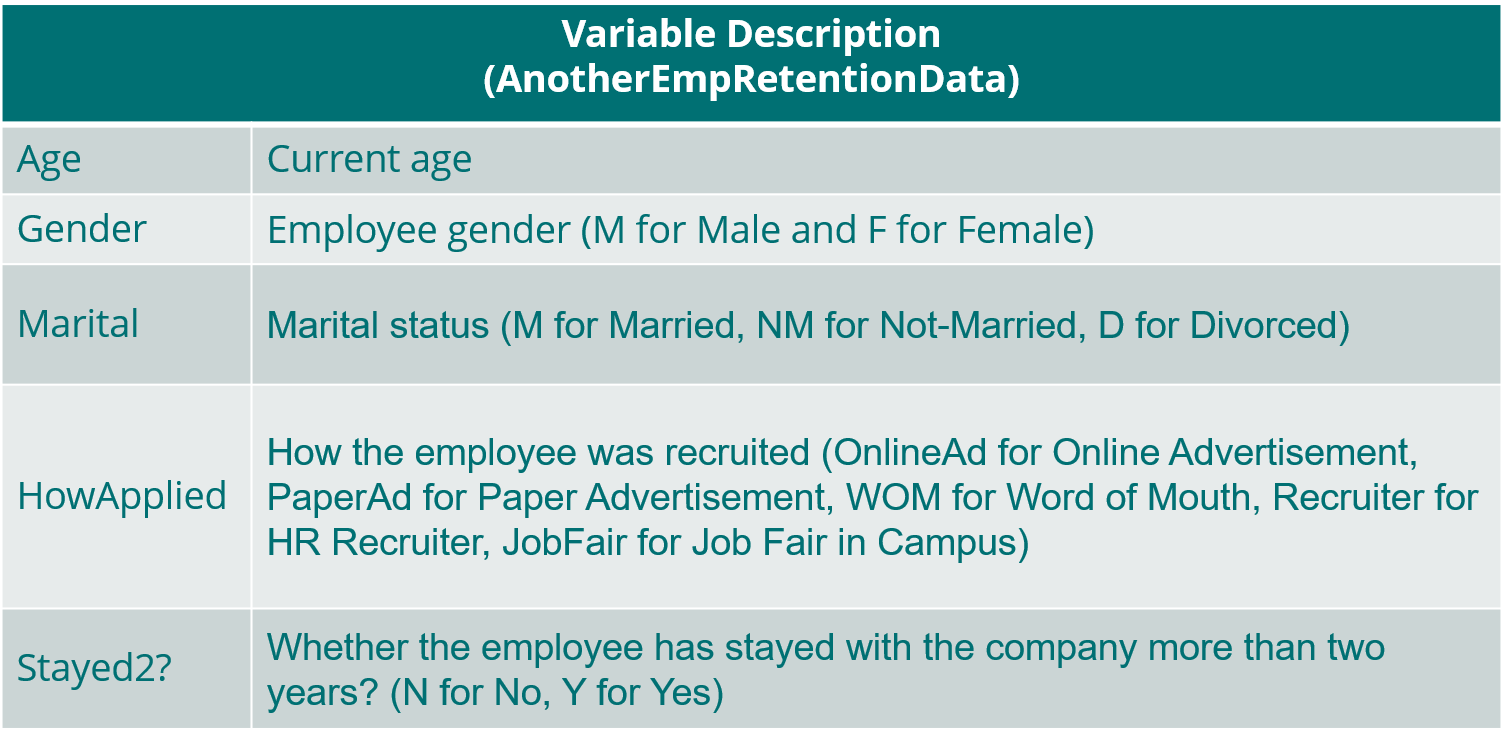
Weighted Avg. 0.845 0.238 0.842 0.845 0.842 0.634 0.870 0.850

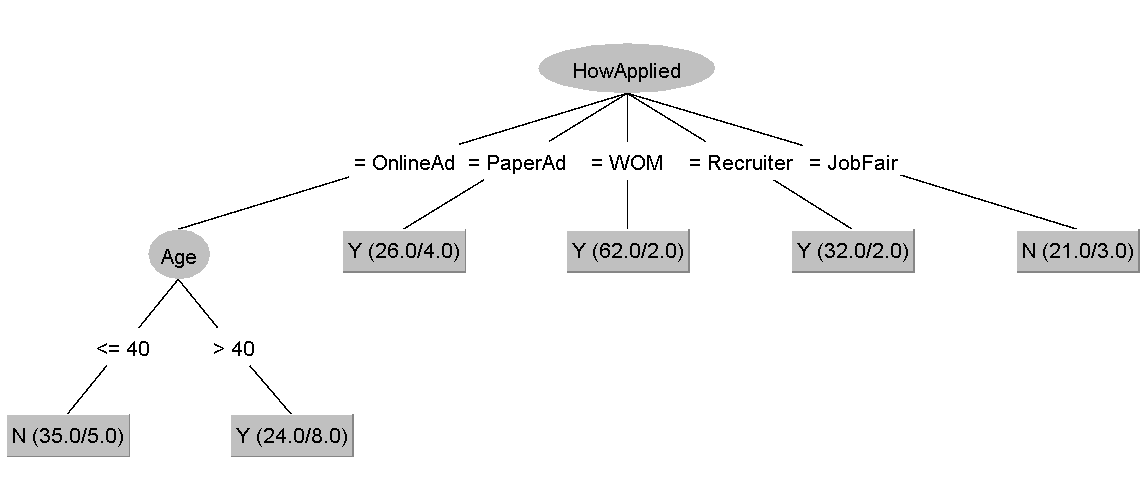
=== Confusion Matrix ===

a b <-- classified as

44 20 | a = N

11 125 | b = Y





1. What variables are the best predictors in terms of which employees are most likely to stay for at least two years?

The classification tree determined that HowApplied and Age were the best predictors for who was most likely to stay for 2 years. However, Age was only relevant if they applied via an online ad.

The gender and marital status variables were deemed irrelevant.

1. Who is most likely to stay for at least two years? Who is most likely to quit within two years? (Not asking for names here; asking for variables)

STAY = OnlineAd applicants over 40, PaperAd applicants, Word-Of-Mouth applicants, and applicants found by a Recruiter.

QUIT = OnlineAd applicants below age 41 and JobFair applicants.

1. What is the overall accuracy of the model? Would you say that this model is fairly accurate, slightly accurate, highly accurate or not very accurate?

This model correctly classified 84.5% of the instances. I would say this model is fairly accurate. You can never have a model with 100% accuracy, but I have created Random Forest models using 14 variables with 90% accuracy. There are clearly more variables at play when it comes to employee turnover.

1. Of the 200 cases, how many did the model predict correctly to stay? How many did the model predict correctly to leave?

The confusion matrix shows 125 Stayers were predicted correctly while 20 were incorrect. 44 Quitters were predicted correctly while 11 were incorrect.

1. Given this information, what advice would you have for Bob about using the most effective ways to recruit loyal employees? (Please provide at least two full sentences)

I would advise Bob to vet applicants under the age of 41 with much more scrutiny. While it is possible those applicants could become loyal employees, chances are they will leave before 2 years are over. Turnover is very costly to organizations and if Bob has a tight budget, applicants under 41 should be “screened out.”

I would also advise Bob to stop wasting time and money on job fairs. The model shows Word-Of-Mouth to be the most effective at recruiting loyal employees so the HR department should reward employees for referring their friends to job roles.