**SEng 474 A1**

1. **Weka classifier output:**

=== Run information ===

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

Relation: titanic2\_1.txt

Instances: 2201

Attributes: 4

pclass

age

sex

survived

Test mode: 10-fold cross-validation

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

J48 pruned tree

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

sex = male

| pclass = 1st

| | age = adult: no (175.0/57.0)

| | age = child: yes (5.0)

| pclass = 2nd

| | age = adult: no (168.0/14.0)

| | age = child: yes (11.0)

| pclass = 3rd: no (510.0/88.0)

| pclass = crew: no (862.0/192.0)

sex = female

| pclass = 1st: yes (145.0/4.0)

| pclass = 2nd: yes (106.0/13.0)

| pclass = 3rd: no (196.0/90.0)

| pclass = crew: yes (23.0/3.0)

Number of Leaves : 10

Size of the tree : 15

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 1737 78.9187 %

Incorrectly Classified Instances 464 21.0813 %

Kappa statistic 0.429

Mean absolute error 0.312

Root mean squared error 0.3959

Relative absolute error 71.3177 %

Root relative squared error 84.6545 %

Total Number of Instances 2201

=== Detailed Accuracy By Class ===

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

0.376 0.013 0.930 0.376 0.535 0.503 0.746 0.680 yes

0.987 0.624 0.768 0.987 0.864 0.503 0.746 0.822 no

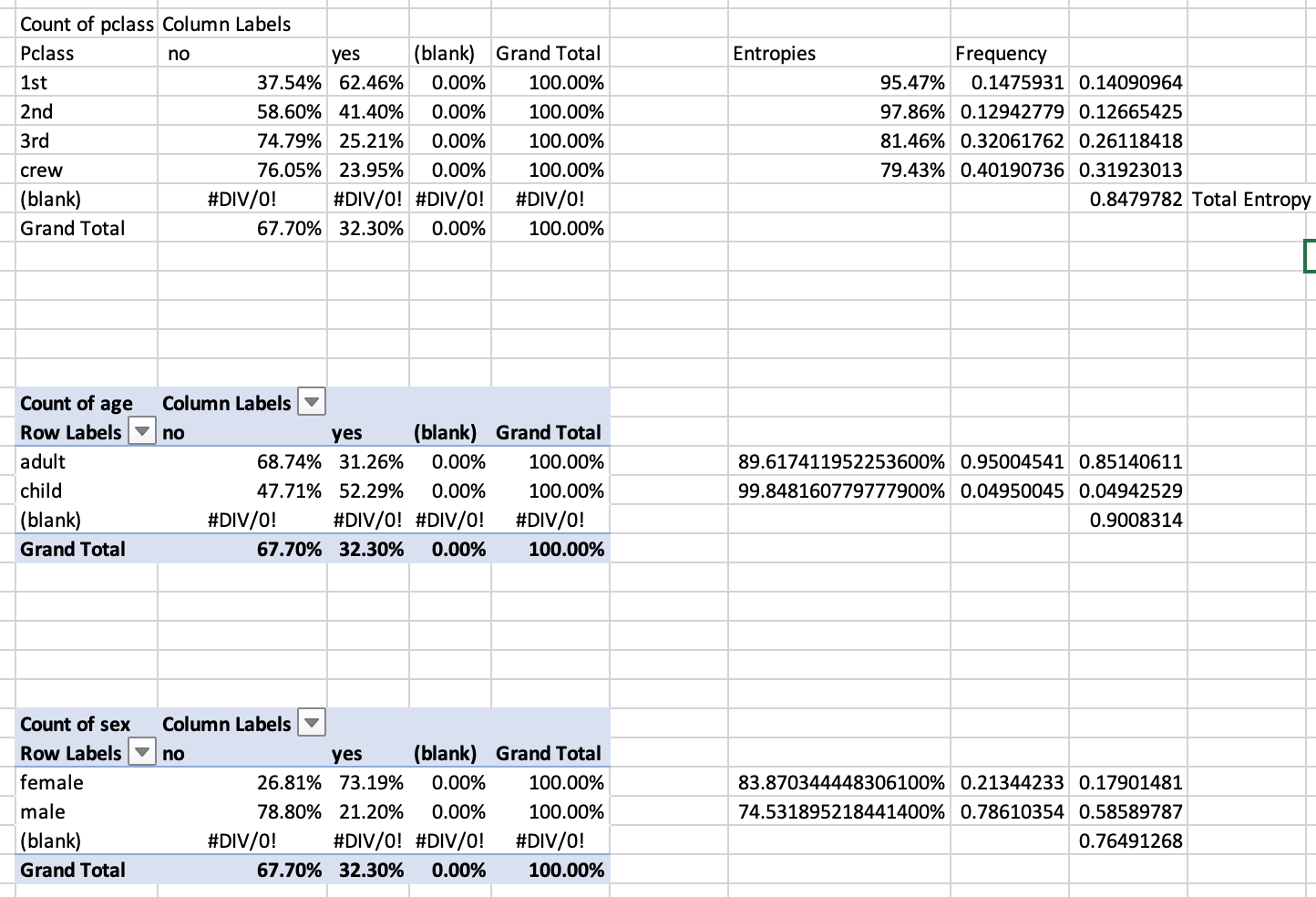
Weighted Avg. 0.789 0.427 0.820 0.789 0.758 0.503 0.746 0.777

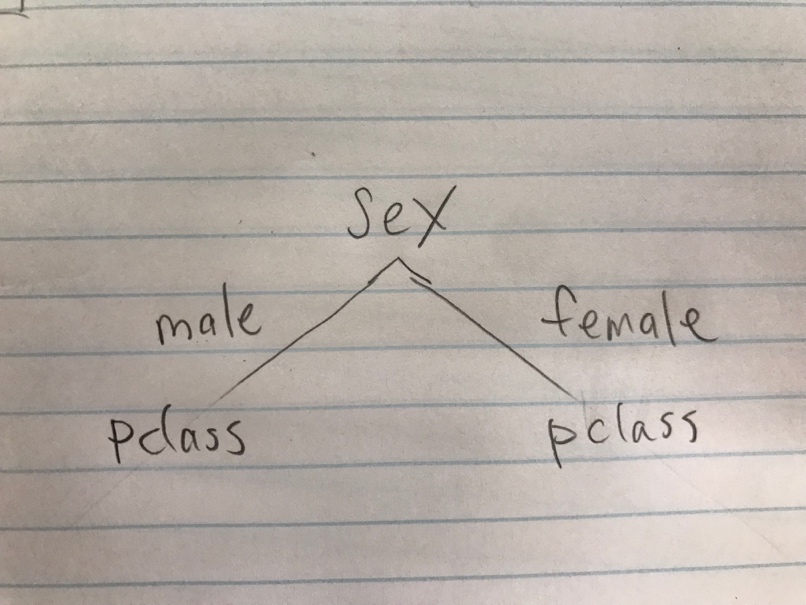
=== Confusion Matrix ===

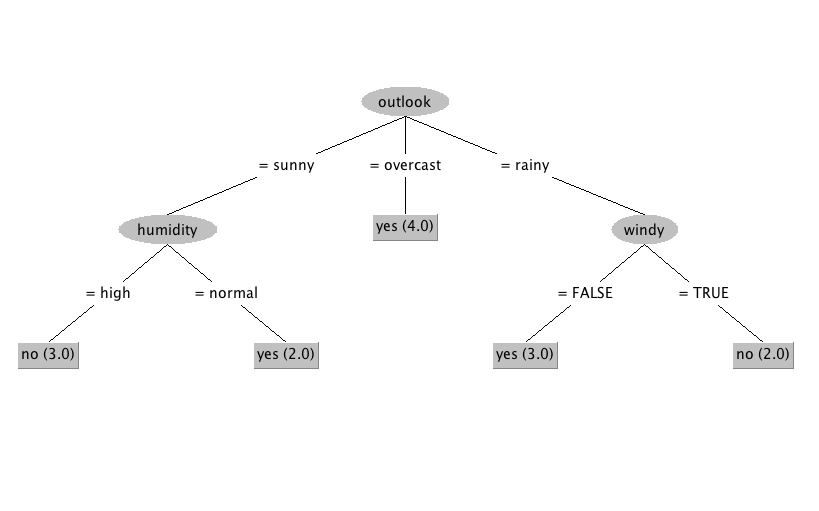
a b <-- classified as

267 444 | a = yes

20 1470 | b = no

**Entropy Calculation**:

**First level decision tree:**

1. **Decision tree built on weather.nominal.csv**

Based on the tree we can follow the pseudo code for PRISM algorithm and obtain the following two:

R1: (outlook = sunny) ∧ (humidity = high) = no

R2: (outlook = rainy) ∧ (windy = false) = yes

**Weka verifications:**

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

Prism rules

----------

If outlook = sunny

and humidity = high then no

If outlook = rainy

and windy = TRUE then no

If outlook = overcast then yes

If humidity = normal

and windy = FALSE then yes

If temperature = mild

and humidity = normal then yes

If outlook = rainy

and windy = FALSE then yes

1. 1. **(2nd, child, male)**

P(survived=yes|E)

=P(Pclass=2nd|survived=yes)\*P(age=child|survived=yes)\*P(sex=male|survived=yes)\*P(survived=yes) / P(E)

= 118/711 \* 57/711 \* 367/711 \* 711/2201 / P(E) = 0.002218 / P(E)

P(survived=no|E)

=P(Pclass=2nd|survived=no)\*P(age=child|survived=no)\*P(sex=male|survived=no)\*P(survived=yes) / P(E)

= 167/1490 \* 52/1490 \* 1364/1490 \* 1490/2201 / P(E) = 0.002424 / P(E)

P(survived=yes|E) = 0.002218 / (0.002218+0.002424) = 0.4778112882 = 47.78%

P(survived=no|E) = 0.002424 / (0.002218+0.002424) = 0.522188712 = 52.22%

* 1. **(2nd, adult, female)**

P(survived=yes|E)

=P(Pclass=2nd|survived=yes)\*P(age=adult|survived=yes)\*P(sex=female|survived=yes)\*P(survived=yes) / P(E)

= 118/711 \* 654/711 \* 344/711 \* 711/2201 / P(E) = 0.02386 / P(E)

P(survived=no|E)

=P(Pclass=2nd|survived= no)\*P(age=adult|survived= no)\*P(sex=female|survived= no)\*P(survived= no) / P(E)

= 167/1490 \* 1438/1490 \* 126/1490 \* 1490/2207 = 0.006175 / P(E)

P(survived=yes|E) = 0.02386 / (0.02386+0.006175) = 79.44%

P(survived=no|E) = 0.006175 / (0.02386+0.006175) = 20.56%

**Weka verification:**

=== Predictions on test set ===

inst# actual predicted error prediction

1 1:? 2:no 0.523

2 1:? 1:yes 0.793