Objective:

First things first the weather data used in this report is sourced from the Kaggle competition data: Predict rain tomorrow in Australia. The data contains several meteorological observations as attributes, and the class attribute “Rain Tomorrow” which will be our focus in this report. Our report aims to

Location MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustSpeed

8.544 5.224 7.003 5.464 3.635 3.966 14.368

WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am

9.594 9.098 18.277 18.452 7.189 7.160 2.864

Cloud3pm Temp9am Temp3pm

2.670 5.901 6.718

Question 1:

Standard Deviations:

Mean:

Question 10

predicted

observed 0 1

0 145 9

1 25 21

Question 5 (Confusion matrix)

Decision Tree

predicted

observed No Yes

No 145 16

Yes 17 22

Naive Bayes

predicted

actual No Yes

No 137 24

Yes 14 25

Bagging

predicted

observed No Yes

No 152 9

Yes 23 16

Boosting

predicted

observed No Yes

No 148 13

Yes 18 21

Random Forest

predicted

observed No Yes

No 153 8

Yes 21 18

Question 6:

AUCs:

Decision Tree:

[1] 0.7626

Naïve Bayes:

[1] 0.8207

Bagging:

[1] 0.8344

Boosting:

[1] 0.8446

Random Forest:

[1] 0.8646

Question 8:

Decision Tree:

Classification tree:

tree(formula = RainTomorrow ~ ., data = WAUSComplete.train)

Variables actually used in tree construction:

[1] "Sunshine" "Pressure3pm" "Humidity3pm" "Month" "Location" "WindSpeed9am"

[7] "Evaporation" "Day" "WindGustSpeed" "MinTemp" "Cloud3pm" "Year"

Number of terminal nodes: 22

Residual mean deviance: 0.341 = 151 / 442

Misclassification error rate: 0.0711 = 33 / 464

Bagging:

Cloud3pm Cloud9am Day Evaporation Humidity3pm Humidity9am Location

0.8871 0.0000 0.9796 2.1676 27.9972 0.0000 0.4311

MaxTemp MinTemp Month Pressure3pm Pressure9am Rainfall RainToday

0.0000 2.8250 1.9633 20.0864 5.1337 2.3536 0.0000

Sunshine Temp3pm Temp9am WindGustSpeed WindSpeed3pm WindSpeed9am Year

17.9392 3.1129 1.4920 6.7174 1.5516 2.6893 1.6730

Boosting:

Cloud3pm Cloud9am Day Evaporation Humidity3pm Humidity9am Location

0.9847 0.9418 7.9044 7.0869 8.6936 3.7528 0.9745

MaxTemp MinTemp Month Pressure3pm Pressure9am Rainfall RainToday

1.0332 6.5685 0.3647 12.7892 6.7399 6.4371 0.0000

Sunshine Temp3pm Temp9am WindGustSpeed WindSpeed3pm WindSpeed9am Year

17.5278 2.5538 3.0164 6.1962 2.0682 3.1792 1.1873

Random Forest:

MeanDecreaseGini

Day 4.981

Month 3.712

Year 3.386

Location 2.322

MinTemp 7.294

MaxTemp 5.407

Rainfall 7.687

Evaporation 7.591

Sunshine 17.940

WindGustSpeed 8.200

WindSpeed9am 4.447

WindSpeed3pm 4.200

Humidity9am 6.375

Humidity3pm 17.944

Pressure9am 13.329

Pressure3pm 17.979

Cloud9am 3.735

Cloud3pm 6.119

Temp9am 6.047

Temp3pm 7.323

RainToday 1.098

Question 9 (pruned decision tree)

predicted

actual No Yes

No 153 8

Yes 22 17

Classification tree:

snip.tree(tree = WAUS.decision.tree, nodes = c(9L, 3L, 17L, 5L

))

Variables actually used in tree construction:

[1] "Sunshine" "Pressure3pm" "Humidity3pm"

Number of terminal nodes: 5

Residual mean deviance: 0.671 = 308 / 459

Misclassification error rate: 0.121 = 56 / 464