Q9:

NB is the model that can be done by hand.

Assumed variables are independent, so some variables are omitted.

To make such a model, the current WAUS training dataset has too many rows, thus we first sample 1% from which to be the new dataset. 70 % of this data set was spilt as a training set (14 rows) while the rest was testing set (6 rows).

A screenshot of a computer

Description automatically generated with medium confidence

Let’s test the first row:

P(Y) = The probilbility of being Cloudy = .4286

P(N) = The probilbility of not being Cloudy = .5714

Row 1

P(Not Humid|Y) = .67; P(Not Cold|Y) = 1; P(Not Sunny|Y) = 1; P(Not Windy|Y) = 1

P(Not Humid|N) = .625; P(Not Cold|N) = .5; P(Not Sunny|N) = 1; P(Not Windy|N) = 1

So, P(Y) \* P(A|Y) = .67 \* 1 \* 1 \* 1 \* .4286 = .2872

So, P(N) \* P(A|N) = .625 \* .5 \* 1 \* 1 \* .4286 = .1786

P(Y) \* P(A|Y) > , P(N) \* P(A|N)

So, cloudy

Row 2

P(Humid|Y) = .33; P(Not Cold|Y) = 1; P(Not Sunny|Y) = 1; P(Not Windy|Y) = 1

P(Humid|N) = .375; P(Not Cold|N) = .5; P(Not Sunny|N) = 1; P(Not Windy|N) = 1

So, P(Y) \* P(A|Y) = .33 \* 1 \* 1 \* 1 \* .4286 = .33

So, P(N) \* P(A|N) = . 375\* .5 \* 1 \* 1 \* .4286 = .1875

P(Y) \* P(A|Y) > , P(N) \* P(A|N)

So, cloudy