*Leo Toueg*

*20062982*

CISC 251 Exercise Sheet 2

1. Using the Naïve Bayes classifier as a predictor, I made the following workflow in KNIME:

Chart

Description automatically generated

Executing that gave us the following classification accuracy and confusion matrix:

Graphical user interface, table

Description automatically generated with medium confidenceGraphical user interface, text, application, chat or text message

Description automatically generated

As we can see, the accuracy of our model was 98%! This is very good. Also, the following table shows some of the things we discussed in class relating to gaussian distributions of the attributes:

Table

Description automatically generated

As we can see, we can individually look at the distribution of an attribute within a class, such as the alkalinity of ash.

**BINNING:**

I had some trouble implementing this one, so I’m not sure if my results are very accurate, however this is the workflow I constructed:

Diagram

Description automatically generated with medium confidence

Using this workflow, here is my confusion matrix and the classification accuracy:

**Confusion Matrix:**

Table

Description automatically generated

**Classification Accuracy:**

Graphical user interface, text, application, chat or text message

Description automatically generated

As we can see, this model using binning performed almost the same as the first one without binning, however slightly better. (98.8% vs 98.3%)

The binning I used was applied to ***every*** attribute.

UPDATE: I figured it out, here is binning with just one attribute (alcohol content):

Table

Description automatically generated(as we can see, alcohol content is binned)

Here are the results:

**Confusion matrix:**

Table

Description automatically generated with medium confidence

**Classification accuracy:**

A picture containing table

Description automatically generated

**97.17%.**

In conclusion, binning one attribute resulted in a worse classification than without binning.

However, binning all attributes resulted in a slightly better model. However, I didn’t separate the data into training and testing data, and this is probably why my numbers were so high. Im not sure whether or not I was supposed to.

The end

:)