

Practical 1: Naïve Bayes Classification with Weka

I. Building the Naïve Bayes classifier.

1. Run Weka, click on “Explorer” mode button.
2. Click on “Open files” tab and select the “weather.nominal.arff” file from the *data* directory
3. Click on “Classify” tab and click on “choose” button. Then select the algorithm “weka/classifiers/bayes/NaiveBayes”.
4. Choose “Use training set” in the test options.
5. Click “Start”.

II. Predicting/classifying new data

Assume that we have a new observation (evidence): outlook=sunny, temperature=Cool, Humidity=High, Windy=TRUE. Using the above built model, we now predict the event “play” based on the evidence. In other words, we want to classify a new instance: {sunny, Cool, High, TRUE, ?}. (The class attribute is ? because we don’t know the classification.) Do the following:

1. Guess a value for ? (or set it at random), say **no**, i.e.

sunny, Cool, High, TRUE, no

2. Create a test file and include the above instance after @data.
 - Copy the attribute definitions from the training ARFF file into a new test ARFF file.
 - Include a proper name for the relation in the test file, say @relation weather-test
 - Include your test data after the @data statement. This may be a single instance (if you want to classify this instance) or a set of instances (if you want to evaluate the classifier).

Example: weather-test.arff

```
@relation weather-test

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {Hot, Mild, Cool}
@attribute humidity {High, Normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny, Cool, High, TRUE, no
```

3. Use the “Supplied test set” option and load your test file.
4. Click “Start” to classify the provided test data. Correctly Classified Instances (shown on top) tells you that your guess was correct (according to Naïve Bayes).

III. Using numerical dataset.

1. Click on “Open files” tab and select the “diabetes.arff” file from the data directory
2. Click on “Classify” tab and click on “choose” button. Then select the algorithm “weka/classifiers/bayes/NaiveBayes”.
3. In the test options, choose Percentage split 75%. This means that we use 75% of the samples (instances) for training the model and using the 25% for testing (evaluating) the model.
4. Click “Start”.
5. Fill in the following table. The “Actual class” is the real class label of the test data. The “Predicted class” is the class result predicted by the model.

	Predicted class: Negative	Predicted class: Positive
Actual class: Negative	TN=	FP=
Actual class: Positive	FN=	TP=

TN: True Negative; FP: False Positive; FN: False Negative; TP: True Positive

Calculate:

$$\text{Specificity} = \text{TN}/(\text{TN}+\text{FP})$$

$$\text{Sensitivity} = \text{TP}/(\text{TP}+\text{FN})$$

6. Discretise the diabetes dataset.
 - Go to Preprocess tab.
 - In the Filter box, choose filters -> supervised -> attribute -> discretise.
 - Use default settings and click apply.
7. With the discretised dataset, build and evaluate the model with 75% of samples for training and 25% of samples for testing.
8. Compare the Specificity and Sensitivity in Step 5 and Step 7.