These are the WEKA printouts for both classifiers, labeled with what the training set was. For Joyce, the test set was always Ulysses, and for each other author the test set is what the training set excluded.

Everything but ulysses

```
=== Run information ===
```

Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W

weka.classifiers.bayes.NaiveBayes

Relation: author_detection

Instances: 8 Attributes: 4

> lexical_diversity unique_words word_count author

Test mode: user supplied test set: size unknown (reading incrementally)

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

01---

InputMappedClassifier:

Naive Bayes Classifier

Attribute	Class Virginia Woolf (0.36)	•	e Arthur Conan Doyle (0.36)	
=====	========	========	==========	==========
lexical_diversity				
mean	0.1496	0.1654	0.2047	
std. dev.	0.0445	0.0039	0.0589	
weight sum	3	2	3	
precision	0.0236	0.0236	0.0236	
unique_words				
mean	16090	12642.1429	6895.7143	
std. dev.	3250.6709	1149.2857	3250.6709	
weight sum	3	2	3	
precision	2298.5714	2298.571	4 2298.5714	

word	count

mean	111081.4286	77757	37027.1429
std. dev.	47992.6623	11108.1429	27708.5765
weight sum	3	2	3
precision	22216.2857	22216.2857	22216.2857

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 %
Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error0.6667Root mean squared error0.8165Relative absolute error137.5 %Root relative squared error158.7713 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

	TP R	ate FP	Rate P	recision	n Reca	II F-	Measure	MCC	ROC Area PRC Area
Class									
	?	1.000	0.000	?	?	?	?	?	Virginia Woolf
	0.000	?	?	0.000	?	?	?	1.000	James Joyce
	?	0.000	?	?	?	?	?	? A	rthur Conan Doyle
Weighted	Avg.	0.000	?	?	0.000	?	?	?	1.000

```
=== Confusion Matrix ===
a b c <-- classified as
000 | a = Virginia Woolf
100|b=James Joyce
0 0 0 | c = Arthur Conan Doyle
Dubliners
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author_detection
Instances: 7
Attributes: 4
       lexical_diversity
       unique_words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
               Virginia Woolf
Attribute
                               James Joyce Arthur Conan Doyle
                 (0.4)
                             (0.2)
                                        (0.4)
______
=====
lexical_diversity
 mean
                    0.1562
                                 0.1654
                                              0.2021
 std. dev.
                    0.0468
                                0.0046
                                             0.0566
weight sum
 precision
                    0.0276
                                 0.0276
                                              0.0276
unique_words
```

10726.6667

446.9444

mean

std. dev.

16983.8889

2528.2996

6257.2222

2528.2996

weight sum	3	1	3
precision	2681.6667	2681.6667	2681.6667
word_count			
mean	112315.6667	77757	34558.6667
std. dev.	44053.8289	4319.8333	24436.6675
weight sum	3	1	3
precision	25919	25919	25919

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 % Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error0.6667Root mean squared error0.8165Relative absolute error125 %Root relative squared error144.3376 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

	TP Ra	ate FP	Rate Pi	recision	Recall	F-Me	easure	MCC	ROC Area PRC Area
Class									
	?	1.000	0.000	?	?	?	?	?	Virginia Woolf
	0.000	?	?	0.000	?	?	?	1.000	James Joyce

```
?
              0.000 ?
                                                         Arthur Conan Doyle
Weighted Avg. 0.000 ?
                           ?
                                 0.000 ?
                                             ?
                                                    ?
                                                          1.000
=== Confusion Matrix ===
a b c <-- classified as
0 0 0 | a = Virginia Woolf
100|b = James Joyce
0 0 0 | c = Arthur Conan Doyle
Only portrait
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author_detection
Instances: 7
Attributes: 4
       lexical_diversity
       unique_words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
Attribute
               Virginia Woolf
                               James Joyce Arthur Conan Doyle
                 (0.4)
                             (0.2)
                                        (0.4)
______
lexical_diversity
 mean
                    0.1562
                                 0.1654
                                              0.2021
 std. dev.
                    0.0468
                                 0.0046
                                              0.0566
 weight sum
                        3
                                   1
 precision
                    0.0276
                                 0.0276
                                              0.0276
unique_words
 mean
                  16983.8889
                                 13408.3333
                                                6257.2222
```

std. dev.	2528.2996	446.9444	2528.2996
weight sum	3	1	3
precision	2681.6667	2681.6667	2681.6667
word_count			
mean	112315.6667	77757	34558.6667
std. dev.	44053.8289	4319.8333	24436.6675
weight sum	3	1	3
precision	25919	25919	25919

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 %
Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error 0.6667

Root mean squared error 0.8165

Relative absolute error 125 %

Root relative squared error 144.3376 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
? 1.000 0.000 ? ? ? ? ? Virginia Woolf

```
0.000 ?
                           0.000 ? ? ?
                     ?
                                                   1.000
                                                           James Joyce
         ?
              0.000
                                ?
                                       ?
                                                  ?
                                                        Arthur Conan Doyle
                                            ?
Weighted Avg. 0.000 ?
                          ?
                                0.000 ? ?
                                                   ?
                                                         1.000
=== Confusion Matrix ===
a b c <-- classified as
000|a = Virginia Woolf
1 0 0 | b = James Joyce
0 0 0 | c = Arthur Conan Doyle
Night and Day
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author_detection
Instances: 8
Attributes: 4
       lexical_diversity
       unique_words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
Attribute
              Virginia Woolf
                               James Joyce Arthur Conan Doyle
                 (0.27)
                            (0.36)
                                        (0.36)
______
lexical_diversity
 mean
                    0.1728
                                0.1656
                                             0.2088
std. dev.
                   0.0432
                                0.0102
                                             0.0539
                        2
                                  3
 weight sum
                                             3
 precision
                    0.0216
                                0.0216
                                             0.0216
```

unique_words			
mean	14913.5714	23861.7143	5965.4286
std. dev.	2982.7143	16872.78	994.2381
weight sum	2	3	3
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	108961.2857	133174.9048	36320.4286
std. dev.	36320.4286	85608.0711	29655.5057
weight sum	2	3	3
precision	36320.4286	36320.4286	36320.4286

Model attributes	Incoming attributes
Model attributes	Incoming attributes

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 100 % Incorrectly Classified Instances 0 0 % Kappa statistic 1 Mean absolute error 0 Root mean squared error 0 Relative absolute error 0.0039 % Root relative squared error 0.0045 % Total Number of Instances 1

=== Detailed Accuracy By Class ===

```
TP Rate FP Rate Precision Recall F-Measure MCC
                                                           ROC Area PRC Area
Class
         1.000 ?
                     1.000
                             1.000 1.000
                                            ?
                                                  ?
                                                        1.000
                                                               Virginia Woolf
         ?
              0.000
                    ?
                            ?
                                                   ?
                                 ?
                                                         James Joyce
              0.000 ?
         ?
                            ?
                                                   ?
                                                         Arthur Conan Doyle
                                                       ?
Weighted Avg.
              1.000 ?
                           1.000
                                   1.000 1.000
                                                  ?
                                                             1.000
=== Confusion Matrix ===
a b c <-- classified as
100 | a = Virginia Woolf
0 0 0 | b = James Joyce
0 0 0 | c = Arthur Conan Doyle
A voyage out
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author detection
Instances: 8
Attributes: 4
       lexical_diversity
       unique_words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
Attribute
               Virginia Woolf
                               James Joyce Arthur Conan Doyle
                 (0.27)
                             (0.36)
                                         (0.36)
______
=====
lexical_diversity
```

mean	0.1654	0.1654	0.2047
std. dev.	0.0472	0.0039	0.0589
weight sum	2	3	3
precision	0.0236	0.0236	0.0236
unique_words			
mean	14913.5714	23861.7143	5965.4286
std. dev.	2982.7143	16872.78	994.2381
weight sum	2	3	3
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	127121.5	133174.9048	36320.4286
std. dev.	54480.6429	85608.0711	29655.5057
weight sum	2	3	3
precision	36320.4286	36320.4286	36320.4286

Model attributes	Incoming attributes
------------------	---------------------

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances	1	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic 1			
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0 %		
Root relative squared error	0 %		

```
Total Number of Instances
```

•

=== Detailed Accuracy By Class ===

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

1.000 ? ? 1.000 1.000 1.000 1.000 Virginia Woolf ? 0.000 ? ? ? ? ? ? James Joyce

? 0.000 ? ? ? ? ? Arthur Conan Doyle

Weighted Avg. 1.000 ? 1.000 1.000 ? ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

100|a=Virginia Woolf

0 0 0 | b = James Joyce

0 0 0 | c = Arthur Conan Doyle

Study in Scarlet

=== Run information ===

Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W

weka.classifiers.bayes.NaiveBayes

Relation: author detection

Instances: 8
Attributes: 4

lexical_diversity unique_words word_count

author

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

Naive Bayes Classifier

Class

Attribute Virginia Woolf James Joyce Arthur Conan Doyle

(0.36) (0.36) (0.27)

=====			
lexical_diversity			
mean	0.1496	0.1654	0.2126
std. dev.	0.0445	0.0039	0.0709
weight sum	3	3	2
precision	0.0236	0.0236	0.0236
unique_words			
mean	15907.8095	23861.7143	5965.4286
std. dev.	2812.13	16872.78	994.2381
weight sum	3	3	2
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	133174.9048	133174.9048	36320.4286
std. dev.	45299.5333	85608.0711	36320.4286
weight sum	3	3	2
precision	36320.4286	36320.4286	36320.4286

Model attributes	Incoming attributes
------------------	---------------------

(numeric) lexical_diversity
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count
 (nominal) author
 --> 1 (numeric) lexical_diversity
 --> 2 (numeric) unique_words
 --> 3 (numeric) word_count
 --> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 1 100 % Incorrectly Classified Instances 0 0 %

Kappa statistic

Mean absolute error 0.0008

```
Root mean squared error
                                 0.0009
Relative absolute error
                               0.1668 %
Root relative squared error
                                 0.1752 %
Total Number of Instances
                                 1
=== Detailed Accuracy By Class ===
         TP Rate FP Rate Precision Recall F-Measure MCC
                                                                ROC Area PRC Area
Class
         ?
               0.000 ?
                              ?
                                    ?
                                                ?
                                                       ?
                                                             Virginia Woolf
          ?
               0.000 ?
                                    ?
                                           ?
                                                 ?
                              ?
                                                             James Joyce
          1.000 ?
                                                      ?
                       1.000
                               1.000 1.000
                                                            1.000
                                                                    Arthur Conan Doyle
                                      1.000 1.000
                                                      ?
                                                            ?
                                                                  1.000
Weighted Avg. 1.000 ?
                             1.000
=== Confusion Matrix ===
a b c <-- classified as
0 0 0 | a = Virginia Woolf
0 0 0 | b = James Joyce
0 0 1 | c = Arthur Conan Doyle
Hounds of Baskervilles
=== Run information ===
Scheme:
            weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
           author_detection
Instances: 8
Attributes: 4
        lexical_diversity
        unique_words
        word_count
        author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
```

Class

Naive Bayes Classifier

Attribute	Virginia Woolf	James Joyce A	rthur Conan Doyle	
	(0.36)	(0.36) (0.2	7)	
=========				
=====				
lexical_diversity				
mean	0.1496	0.1654	0.2362	
std. dev.	0.0445	0.0039	0.0472	
weight sum	3	3	2	
precision	0.0236	0.0236	0.0236	
unique_words				
mean	15907.8095	23861.7143	5965.4286	
std. dev.	2812.13	16872.78	994.2381	
weight sum	3	3	2	
precision	5965.4286	5965.4286	5965.4286	
word_count				
mean	133174.9048	133174.9048	18160.2143	
std. dev.	45299.5333	85608.0711	18160.2143	
weight sum	3	3	2	
precision	36320.4286	36320.4286	36320.4286	

Model attributes	Incoming attributes		
(numeric) lexical_diversity	> 1 (numeric) lexical_dive		

(numeric) lexical_diversity
(numeric) unique_words
(numeric) word_count
(nominal) author
--> 1 (numeric) lexical_diversity
--> 2 (numeric) unique_words
--> 3 (numeric) word_count
--> 4 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 1 100 % Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error0.1964Root mean squared error0.2324Relative absolute error40.5131 %Root relative squared error45.1875 %

Total Number of Instances

=== Detailed Accuracy By Class ===

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

? 0.000 ? ? ? ? ? ? Yirginia Woolf
 ? 0.000 ? ? ? ? ? ? James Joyce

1.000 ? 1.000 1.000 ? ? 1.000 Arthur Conan Doyle

Weighted Avg. 1.000 ? 1.000 1.000 ? ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

0 0 0 | a = Virginia Woolf

0 0 0 | b = James Joyce

0 0 1 | c = Arthur Conan Doyle

^ Old Classifier — New Classifier v

onDubliners

=== Run information ===

Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W

weka.classifiers.bayes.NaiveBayes

Relation: author_detection

Instances: 7

Attributes: 7

avSentenceLength avWordLength lexical_diversity numPronouns unique_words word_count author

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

Naive Bayes Classifier

000
1055

Attribute	Virginia Woolf (0.4)	James Joyce (0.2) (0.4	Arthur Conan Doyle 4)	
=========	:========	=========	, ===========	=======================================
====				
avSentenceLeng	ıth			
mean	83.5467	70.833	76.2817	
std. dev.	11.196	0.9081	8.8977	
weight sum	3	1	3	
precision	5.4487	5.4487	5.4487	
avWordLength				
mean	4.699	4.5734	4.5375	
std. dev.	0.0915	0.009	0.0671	
weight sum	3	1	3	
precision	0.0538	0.0538	0.0538	
lexical_diversity				
mean	0.1562	0.1654	0.2021	
std. dev.	0.0468	0.0046	0.0566	
weight sum	3	1	3	
precision	0.0276	0.0276	0.0276	
numPronouns				
mean	224631.3333		69117.3333	
std. dev.	88107.6578	8639.6667	48873.3351	
weight sum	3	1	3	

precision	51838	51838	51838
unique_words			
mean	16983.8889	10726.6667	6257.2222
std. dev.	2528.2996	446.9444	2528.2996
weight sum	3	1	3
precision	2681.6667	2681.6667	2681.6667
word_count			
mean	112315.6667	77757	34558.6667
std. dev.	44053.8289	4319.8333	24436.6675
weight sum	3	1	3
precision	25919	25919	25919

Model attributes	Incoming attributes
------------------	---------------------

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

(numeric) avWordLength
 (numeric) lexical_diversity
 (numeric) numPronouns
 (numeric) numPronouns
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 % Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error 0.6667

Root mean squared error 0.8165

Relative absolute error 125 %

Root relative squared error 144.3376 %

```
Total Number of Instances
=== Detailed Accuracy By Class ===
        TP Rate FP Rate Precision Recall F-Measure MCC
                                                       ROC Area PRC Area
Class
                           ?
                                                 ?
             1.000 0.000
                                                      Virginia Woolf
        0.000 ?
                    ?
                          0.000 ?
                                      ?
                                           ?
                                                 1.000 James Joyce
             0.000 ?
                          ?
                               ?
                                               ?
                                                     Arthur Conan Doyle
                                     ?
                                          ?
Weighted Avg. 0.000 ?
                         ?
                               0.000 ? ?
                                                ?
                                                      1.000
=== Confusion Matrix ===
a b c <-- classified as
000|a = Virginia Woolf
100|b=James Joyce
0 0 0 | c = Arthur Conan Doyle
onPortrait
=== Run information ===
          weka.classifiers.misc.InputMappedClassifier -I -trim -W
Scheme:
weka.classifiers.bayes.NaiveBayes
Relation:
         author detection
Instances: 7
Attributes: 7
      avSentenceLength
       avWordLength
       lexical diversity
       numPronouns
       unique_words
      word_count
      author
Test mode: user supplied test set: size unknown (reading incrementally)
```

Naive Bayes Classifier

InputMappedClassifier:

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

Class Attribute Virginia Woolf James Joyce Arthur Conan Doyle (0.4)(0.2)(0.4)avSentenceLength mean 83.5467 81.7304 76.2817 11.196 0.9081 8.8977 std. dev. weight sum 3 1 3 precision 5.4487 5.4487 5.4487 avWordLength 4.699 mean 4.5196 4.5375 std. dev. 0.0915 0.009 0.0671 weight sum 3 1 3 0.0538 precision 0.0538 0.0538 lexical_diversity 0.1562 0.1654 0.2021 mean std. dev. 0.0468 0.0046 0.0566 weight sum 3 1 3 precision 0.0276 0.0276 0.0276 numPronouns mean 224631.3333 155514 69117.3333 std. dev. 88107.6578 8639.6667 48873.3351 3 weight sum 3 1 precision 51838 51838 51838 unique words 13408.3333 mean 16983.8889 6257.2222 std. dev. 2528.2996 446.9444 2528.2996 weight sum 3 1 3 2681.6667 precision 2681.6667 2681.6667 word_count mean 112315.6667 77757 34558.6667 std. dev. 44053.8289 24436.6675 4319.8333 weight sum 3 1 3

Attribute mappings:

25919

25919

25919

precision

Model	attributes
woder	allibules

Incoming attributes

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

(numeric) avWordLength
 (numeric) lexical_diversity
 (numeric) numPronouns
 (numeric) unique_words
 --> 2 (numeric) avWordLength
 --> 3 (numeric) lexical_diversity
 --> 4 (numeric) numPronouns
 --> 5 (numeric) unique_words

(numeric) word_count --> 6 (numeric) word_count

(nominal) author --> 7 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 %
Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error 0.6667

Root mean squared error 0.8165

Relative absolute error 125 %

Root relative squared error 144.3376 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

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

Class

? 1.000 0.000 ? ? ? Virginia Woolf 0.000 ? ? 0.000 ? 1.000 James Joyce ? ? ? ? 0.000 ? ? ? Arthur Conan Doyle ? Weighted Avg. 0.000 ? ? 0.000 ? ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

000 | a = Virginia Woolf

1 0 0 | b = James Joyce


```
onPortraitandDubliners
=== Run information ===
            weka.classifiers.misc.InputMappedClassifier -I -trim -W
Scheme:
weka.classifiers.bayes.NaiveBayes
          author_detection
Relation:
Instances: 8
Attributes: 7
        avSentenceLength
        avWordLength
        lexical_diversity
        numPronouns
        unique_words
        word_count
        author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
```

InputMappedClassifier:

Naive Bayes Classifier

Attribute	Class Virginia Woolf (0.36)	James Joy (0.27)	ce Arthur Conan Doyle (0.36)	
====				
avSentenceLeng	jth			
mean	82.5088	77.0601	74.725	
std. dev.	7.938	7.0055	10.089	
weight sum	3	2	3	
precision	4.6703	4.6703	4.6703	
avWordLength	4 7404	4 5406	4 525	
mean	4.7194	4.5196	4.535	

std. dev.	0.0784	0.0461	0.0575
weight sum	3	2	3
precision	0.0461	0.0461	0.0461
lexical_diversity			
mean	0.1496	0.1654	0.2047
std. dev.	0.0445	0.0039	0.0589
weight sum	3	2	3
precision	0.0236	0.0236	0.0236
numPronouns			
	222462 0574	155511	74054 2057
mean	222162.8571	155514	74054.2857
std. dev.	95985.3247	22216.2857	55417.153
weight sum	3	2	3
precision	44432.5714	44432.5714	44432.5714
unique_words			
mean	16090	12642.1429	6895.7143
std. dev.	3250.6709	1149.2857	3250.6709
weight sum	3	2	3
precision	2298.5714	2298.5714	2298.5714
•			
word_count			
mean	111081.4286	77757	37027.1429
std. dev.	47992.6623	11108.1429	27708.5765
weight sum	3	2	3
precision	22216.2857	22216.2857	22216.2857

Model attributes	Incoming attributes
(numeric) avSentenceLength	> 1 (numeric) avSentenceLength
(numeric) avWordLength	> 2 (numeric) avWordLength
(numeric) lexical_diversity	> 3 (numeric) lexical_diversity
(numeric) numPronouns	> 4 (numeric) numPronouns
(numeric) unique_words	> 5 (numeric) unique_words
(numeric) word_count	> 6 (numeric) word_count
(nominal) author	> 7 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 %
Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error 0.6667

Root mean squared error 0.8165

Relative absolute error 137.5 %

Root relative squared error 158.7713 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

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

? 1.000 0.000 ? ? ? Virginia Woolf 0.000 ? ? 0.000 1.000 James Joyce ? 0.000 ? ? ? ? ? ? Arthur Conan Doyle

Weighted Avg. 0.000 ? ? 0.000 ? ? ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

000 | a = Virginia Woolf

100|b = James Joyce

0 0 0 | c = Arthur Conan Doyle

onScarlet

=== Run information ===

Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W

weka.classifiers.bayes.NaiveBayes

Relation: author detection

Instances: 8

Attributes: 7

avSentenceLength avWordLength lexical_diversity numPronouns unique_words word_count author

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

Naive Bayes Classifier

\sim	_	~~
	-	~

Attribute	Virginia Woolf	James Joyo	ce Arthur Conan Doyle	
	(0.36)	(0.36)	(0.27)	
=========	========	========		============
=====				
avSentenceLeng	ıth			
mean	83.4066	70.7143	70.7143	
std. dev.	11.1772	8.8828	5.4396	
weight sum	3	3	2	
precision	5.4396	5.4396	5.4396	
avWordLength				
mean	4.7194	4.5965	4.4965	
std. dev.	0.0784	0.115	0.0231	
weight sum	3	3	2	
precision	0.0461	0.0461	0.0461	
lexical_diversity				
mean	0.1496	0.1654	0.2126	
std. dev.	0.0445	0.0039	0.0709	
weight sum	3	3	2	
precision	0.0236	0.0236	0.0236	
numPronouns				
mean	266349.809	5 266349.8	72640.8571	
std. dev.	90599.0666	171216.14	123 72640.8571	
weight sum	3	3	2	

precision	72640.8571	72640.8571	72640.8571
unique_words			
mean	15907.8095	23861.7143	5965.4286
std. dev.	2812.13	16872.78	994.2381
weight sum	3	3	2
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	133174.9048	133174.9048	36320.4286
std. dev.	45299.5333	85608.0711	36320.4286
weight sum	3	3	2
precision	36320.4286	36320.4286	36320.4286

Model attributes Incoming attributes

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

(numeric) avWordLength
 (numeric) lexical_diversity
 (numeric) numPronouns
 (numeric) numPronouns
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 0 0 %
Incorrectly Classified Instances 1 100 %

Kappa statistic 0

Mean absolute error 0.6077

Root mean squared error 0.6732

Relative absolute error 125.3407 %

Root relative squared error 130.9132 %

```
Total Number of Instances
=== Detailed Accuracy By Class ===
        TP Rate FP Rate Precision Recall F-Measure MCC
                                                       ROC Area PRC Area
Class
             1.000 0.000
                                                      Virginia Woolf
        ?
             0.000 ?
                               ?
                                          ?
                                               ?
                                                     James Joyce
        0.000 ?
                          0.000 ?
                                     ?
                                                        Arthur Conan Doyle
                                           ?
                                                 1.000
                                          ?
Weighted Avg. 0.000 ?
                               0.000 ?
                         ?
                                                ?
                                                      1.000
=== Confusion Matrix ===
a b c <-- classified as
000|a = Virginia Woolf
0 0 0 | b = James Joyce
1 0 0 | c = Arthur Conan Doyle
onBrucePartington
=== Run information ===
Scheme:
          weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
         author_detection
Instances: 8
Attributes: 7
      avSentenceLength
       avWordLength
       lexical diversity
       numPronouns
       unique_words
      word_count
```

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

author

Naive Bayes Classifier

Attribute	Class Virginia Woolf (0.36)	James Joyce A (0.36) (0.2	rthur Conan Doyle 7)	
====				
avSentenceLeng	ıth			
mean	83.4066	70.7143	81.5934	
std. dev.	11.1772	8.8828	5.4396	
weight sum	3	3	2	
precision	5.4396	5.4396	5.4396	
avWordLength				
mean	4.7194	4.5965	4.5427	
std. dev.	0.0784	0.115	0.0692	
weight sum	3	3	2	
precision	0.0461	0.0461	0.0461	
lexical_diversity				
mean	0.1574	0.1625	0.1752	
std. dev.	0.0503	0.0072	0.0229	
weight sum	3	3	2	
precision	0.0152	0.0152	0.0152	
numPronouns				
mean	231596.0952	273704.4762	94743.8571	
std. dev.	78777.5673	165780.8616	31581.2857	
weight sum	3	3	2	
precision	63162.5714	63162.5714	63162.5714	
unique_words				
mean	15452.1429	24036.6667	10301.4286	
std. dev.	4205.5406	15921.9198	858.4524	
weight sum	3	3	2	
precision	5150.7143	5150.7143	5150.7143	
word_count				
mean	115798.0476	136852.2381	47371.9286	
std. dev.	39388.7837	82890.4308	15790.6429	
weight sum	3	3	2	
precision	31581.2857	31581.2857	31581.2857	

```
Model attributes
                              Incoming attributes
(numeric) avSentenceLength
                                  --> 1 (numeric) avSentenceLength
(numeric) avWordLength
                           --> 2 (numeric) avWordLength
(numeric) lexical diversity
                           --> 3 (numeric) lexical diversity
                           --> 4 (numeric) numPronouns
(numeric) numPronouns
(numeric) unique_words
                           --> 5 (numeric) unique_words
(numeric) word_count
                           --> 6 (numeric) word count
(nominal) author
                           --> 7 (nominal) author
Time taken to build model: 0 seconds
=== Evaluation on test set ===
Time taken to test model on supplied test set: 0 seconds
=== Summary ===
                                                 %
Correctly Classified Instances
                                  0
Incorrectly Classified Instances
                                           100
                                  1
                                                  %
Kappa statistic
Mean absolute error
                               0.6624
Root mean squared error
                                  0.8113
Relative absolute error
                              136.6246 %
Root relative squared error
                                157.7605 %
Total Number of Instances
                                  1
=== Detailed Accuracy By Class ===
         TP Rate FP Rate Precision Recall F-Measure MCC
                                                                ROC Area PRC Area
Class
          ?
               1.000 0.000
                                                         ?
                                                               Virginia Woolf
          ?
                                    ?
                                                 ?
                                                       ?
               0.000 ?
                                                              James Joyce
         0.000 ?
                              0.000
                                                         1.000
                                                                 Arthur Conan Doyle
Weighted Avg. 0.000 ?
                             ?
                                    0.000 ?
                                                   ?
                                                        ?
                                                               1.000
=== Confusion Matrix ===
```

a b c <-- classified as

```
0 0 0 | b = James Joyce
100 | c = Arthur Conan Doyle
onHounds
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author_detection
Instances: 8
Attributes: 7
       avSentenceLength
       avWordLength
       lexical_diversity
       numPronouns
       unique_words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
               Virginia Woolf James Joyce Arthur Conan Doyle (0.27) (0.36) (0.36)
Attribute
```

	(0.27)	(0.36)	(0.36)	
==========	=======			
=====				
avSentenceLength	1			
mean	89.7528	70.7143	76.1539	
std. dev.	8.1593	8.8828	8.8828	
weight sum	2	3	3	
precision	5.4396	5.4396	5.4396	
•				

avWordLength

000|a = Virginia Woolf

mean	4.7041	4.5965	4.535
std. dev.	0.0922	0.115	0.0575
weight sum	2	3	3
precision	0.0461	0.0461	0.0461
lexical_diversity			
mean	0.1181	0.1654	0.2047
std. dev.	0.0039	0.0039	0.0589
weight sum	2	3	3
precision	0.0236	0.0236	0.0236
numPronouns			
mean	326883.8571	266349.8095	72640.8571
std. dev.	36320.4286	171216.1423	59311.0115
weight sum	2	3	3
precision	72640.8571	72640.8571	72640.8571
unique_words			
mean	17896.2857	23861.7143	5965.4286
std. dev.	994.2381	16872.78	994.2381
weight sum	2	3	3
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	163441.9286	133174.9048	36320.4286
std. dev.	18160.2143	85608.0711	29655.5057
weight sum	2	3	3
precision	36320.4286	36320.4286	36320.4286

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

(numeric) avWordLength
(numeric) lexical_diversity
(numeric) numPronouns
(numeric) unique_words
(numeric) word_count

--> 2 (numeric) avWordLength
--> 3 (numeric) lexical_diversity
--> 4 (numeric) numPronouns
--> 5 (numeric) unique_words
--> 6 (numeric) word_count

(nominal) author --> 7 (nominal) author

```
Time taken to build model: 0 seconds
=== Evaluation on test set ===
Time taken to test model on supplied test set: 0 seconds
=== Summary ===
Correctly Classified Instances
                                       100
                                             %
Incorrectly Classified Instances
                               0
                                        0
                                             %
Kappa statistic
                          1
Mean absolute error
                            0
                              0
Root mean squared error
Relative absolute error
                            0
Root relative squared error
                              0
                                  %
Total Number of Instances
=== Detailed Accuracy By Class ===
         TP Rate FP Rate Precision Recall F-Measure MCC
                                                          ROC Area PRC Area
Class
         ?
                                                  ?
              0.000 ?
                                                        Virginia Woolf
         ?
              0.000 ?
                                ?
                                       ?
                                                        James Joyce
         1.000 ?
                             1.000 1.000 ?
                                                 ?
                                                              Arthur Conan Doyle
                     1.000
                                                       1.000
                                  1.000 1.000 ?
                                                       ?
Weighted Avg. 1.000 ?
                          1.000
                                                             1.000
=== Confusion Matrix ===
a b c <-- classified as
000 | a = Virginia Woolf
000|b = James Joyce
0 0 1 | c = Arthur Conan Doyle
onJacob'sRoom
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
         author detection
Instances: 8
```

Attributes: 7

avSentenceLength

avWordLength lexical_diversity numPronouns unique_words word_count author

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

Naive Bayes Classifier

C	lass
_	las

Attribute	Virginia Woolf (0.27)	James Joyce Ai (0.36) (0.36	thur Conan Doyle 6)
=========	========	=========	
=====			
avSentenceLeng			
mean	75.7339	70.4193	75.7339
std. dev.	3.986	11.4296	8.6107
weight sum	2	3	3
precision	3.986	3.986	3.986
avWordLength			
mean	4.6597	4.5934	4.527
std. dev.	0.0797	0.1045	0.0677
weight sum	2	3	3
precision	0.0398	0.0398	0.0398
lexical_diversity			
mean	0.1728	0.1656	0.2088
std. dev.	0.0432	0.0102	0.0539
weight sum	2	3	3
precision	0.0216	0.0216	0.0216
numPronouns			
mean	217922.571	4 266349.8095	72640.8571
std. dev.	72640.8571	171216.1423	59311.0115
weight sum	2	3	3
precision	72640.8571	72640.8571	72640.8571

14913.5714	23861.7143	5965.4286
2982.7143	16872.78	994.2381
2	3	3
5965.4286	5965.4286	5965.4286
108961.2857	133174.9048	36320.4286
36320.4286	85608.0711	29655.5057
2	3	3
36320.4286	36320.4286	36320.4286
	2982.7143 2 5965.4286 108961.2857 36320.4286 2	2982.7143 16872.78 2 3 5965.4286 5965.4286 108961.2857 133174.9048 36320.4286 85608.0711 2 3

Model attributes Incoming attributes

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

--> 2 (numeric) avWordLength (numeric) avWordLength --> 3 (numeric) lexical diversity (numeric) lexical_diversity (numeric) numPronouns --> 4 (numeric) numPronouns (numeric) unique_words --> 5 (numeric) unique_words (numeric) word_count --> 6 (numeric) word_count (nominal) author --> 7 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 100 % % Incorrectly Classified Instances 0 0

Kappa statistic 1

Mean absolute error 0.0714 Root mean squared error 0.0875 Relative absolute error 14.7274 % 17.0058 % Root relative squared error 1

Total Number of Instances

```
=== Detailed Accuracy By Class ===
        TP Rate FP Rate Precision Recall F-Measure MCC
                                                          ROC Area PRC Area
Class
         1.000 ?
                     1.000
                            1.000 1.000
                                                 ?
                                                      1.000
                                                             Virginia Woolf
         ?
                           ?
              0.000 ?
                                                       James Joyce
              0.000 ?
                           ?
                                                       Arthur Conan Doyle
                                                 ?
                                                      ?
                                                            1.000
Weighted Avg. 1.000 ?
                          1.000
                                  1.000 1.000
=== Confusion Matrix ===
a b c <-- classified as
100| a = Virginia Woolf
0 0 0 | b = James Joyce
0 0 0 | c = Arthur Conan Doyle
onVoyage
=== Run information ===
Scheme:
           weka.classifiers.misc.InputMappedClassifier -I -trim -W
weka.classifiers.bayes.NaiveBayes
Relation:
          author detection
Instances: 8
Attributes: 7
       avSentenceLength
       avWordLength
       lexical_diversity
       numPronouns
       unique words
       word_count
       author
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
                 Class
```

James Joyce Arthur Conan Doyle

(0.36)

Attribute

Virginia Woolf

(0.36)

(0.27)

=====			
avSentenceLength			
mean	84.3132	70.7143	76.1539
std. dev.	13.5989	8.8828	8.8828
weight sum	2	3	3
precision	5.4396	5.4396	5.4396
avWordLength			
mean	4.7733	4.5965	4.535
std. dev.	0.0231	0.115	0.0575
weight sum	2	3	3
precision	0.0461	0.0461	0.0461
Tanaka at Taka a wakina			
lexical_diversity	0.4054	0.4054	0.0047
mean	0.1654	0.1654	0.2047
std. dev.	0.0472	0.0039	0.0589
weight sum	2	3	3
precision	0.0236	0.0236	0.0236
numPronouns			
mean	254243	266349.8095	72640.8571
std. dev.	108961.2857	171216.1423	59311.0115
weight sum	2	3	3
precision	72640.8571	72640.8571	72640.8571
•			
unique_words			
mean	14913.5714	23861.7143	5965.4286
std. dev.	2982.7143	16872.78	994.2381
weight sum	2	3	3
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	127121.5	133174.9048	36320.4286
std. dev.	54480.6429	85608.0711	29655.5057
weight sum	2	3	3
precision	36320.4286	36320.4286	36320.4286
ρισσιοίτι	30320.4200	JUJZU.4200	JUJZU.4Z00

Attribute mappings:

Model attributes	Incoming attributes
Model attributes	Incoming attributes

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

(numeric) avWordLength
 (numeric) lexical_diversity
 (numeric) numPronouns
 (numeric) numPronouns
 (numeric) unique_words
 (numeric) word_count
 2 (numeric) avWordLength
 --> 3 (numeric) lexical_diversity
 --> 4 (numeric) numPronouns
 --> 5 (numeric) unique_words
 --> 6 (numeric) word_count

(nominal) author --> 7 (nominal) author

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 1 100 % Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 1

=== Detailed Accuracy By Class ===

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

? 1.000 ? 1.000 1.000 1.000 ? 1.000 Virginia Woolf ? ? ? 0.000 ? ? James Joyce ? ? ? ? ? 0.000 ? ? Arthur Conan Doyle Weighted Avg. 1.000 ? 1.000 1.000 1.000 ? ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

100|a=Virginia Woolf

0 0 0 | b = James Joyce

0 0 0 | c = Arthur Conan Doyle


```
onNightandDay
=== Run information ===
```

Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W

weka.classifiers.bayes.NaiveBayes

Relation: author_detection

Instances: 8
Attributes: 7

avSentenceLength avWordLength lexical_diversity numPronouns unique_words word_count author

Test mode: user supplied test set: size unknown (reading incrementally)

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

InputMappedClassifier:

Naive Bayes Classifier

Class

Attribute Virginia Woolf James Joyce Arthur Conan Doyle

	(0.27)	(0.36)	(0.36)	
==========	=======	========		===
=====				
avSentenceLength				
mean	75.7339	70.4193	75.7339	
std. dev.	3.986	11.4296	8.6107	
weight sum	2	3	3	
precision	3.986	3.986	3.986	
avWordLength				
mean	4.6597	4.5934	4.527	
std. dev.	0.0797	0.1045	0.0677	
weight sum	2	3	3	
precision	0.0398	0.0398	0.0398	

lexical_diversity			
mean	0.1728	0.1656	0.2088
std. dev.	0.0432	0.0102	0.0539
weight sum	2	3	3
precision	0.0216	0.0216	0.0216
numPronouns			
mean	217922.5714	266349.8095	72640.8571
std. dev.	72640.8571	171216.1423	59311.0115
weight sum	2	3	3
precision	72640.8571	72640.8571	72640.8571
unique_words			
mean	14913.5714	23861.7143	5965.4286
std. dev.	2982.7143	16872.78	994.2381
weight sum	2	3	3
precision	5965.4286	5965.4286	5965.4286
word_count			
mean	108961.2857	133174.9048	36320.4286
std. dev.	36320.4286	85608.0711	29655.5057
weight sum	2	3	3
precision	36320.4286	36320.4286	36320.4286

Model attributes

	•	

(numeric) avSentenceLength --> 1 (numeric) avSentenceLength

Incoming attributes

(numeric) avWordLength
 (numeric) lexical_diversity
 (numeric) numPronouns
 (numeric) unique_words
 (numeric) word_count
 (numeric) word_count

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 1 100 % Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error0.0714Root mean squared error0.0875Relative absolute error14.7274 %Root relative squared error17.0058 %

Total Number of Instances 1

=== Detailed Accuracy By Class ===

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

1.000 ? 1.000 1.000 1.000 ? ? 1.000 Virginia Woolf ? ? 0.000 ? ? ? ? James Joyce ? 0.000 ? Arthur Conan Doyle ? Weighted Avg. 1.000 ? 1.000 1.000 1.000 ? 1.000

=== Confusion Matrix ===

a b c <-- classified as

100 | a = Virginia Woolf

0 0 0 | b = James Joyce

0 0 0 | c = Arthur Conan Doyle