

Result

Maxime Daigle

maxime.daigle.1@umontreal.ca

2 Mai 2018

Contents	
Description du programme d'extraction de caractéristiques.....	1
Naives Bayes	1
Expérience 1	1
Expérience 2	1
Expérience 3	1
Expérience 4	2
Expérience 5	2
Expérience 6	3
Expérience 7	3
Expérience 8	3
Expérience 9	4
Analyse	4
Arbre de décision (J48).....	5
Expérience 1	5
Expérience 2	5
Expérience 3	5
Expérience 4	6
Expérience 5	6
Expérience 6	7
Expérience 7	7
Analyse	7
SVM (SMO)	8
Expérience 1	8
Expérience 2	8
Expérience 3	8
Expérience 4	9
Expérience 5	9
Expérience 6	10
Expérience 7	10
Analyse	10
MultiLayerPerceptron.....	11
Expérience 1	11
Expérience 2	11
Expérience 3	11
Expérience 4	12
Expérience 5	12
Analyse	13
Comparaison entre algorithmes	13

Description du programme d'extraction de caractéristiques

Le programme enlève les ponctuations et, pour chaque instance, prend en note le sens du mot intérêt, les mots dans la fenêtre de contexte et leur catégorie. De plus, le programme a la variable window pour modifier la grandeur de la fenêtre et la variable booléenne use_stoplist pour activer ou désactiver l'élimination des stopwords à l'aide du fichier stoplist-english.txt.

Naives Bayes

Expérience 1

La taille de la fenêtre de contexte est de 2 mots et le fichier stoplist-english.txt est utiliser pour enlever les mot outils. Dans le preprocess de Weka, IDfTransform, TFTransform et lowerCaseToken sont à True. De plus, SnowballStemmer et MultiStopwords sont utilisés.

```
=== Summary ===
Correctly Classified Instances      1753      74.0287 %
Incorrectly Classified Instances    615      25.9713 %
Kappa statistic                    0.5898
Mean absolute error                0.0948
Root mean squared error            0.2444
Relative absolute error            43.9558 %
Root relative squared error        74.4643 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
195    0    6   39   31   90 |  a = 1
  1    0    0    1    4    5 |  b = 2
   3    0   30   12    7   14 |  c = 3
  21    0    4  105   14   34 |  d = 4
  67    0   13   27  308   85 |  e = 5
  45    0    4   21   67 1115 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.540    0.068    0.587    0.540    0.563    0.489    0.901    0.644    1
0.000    0.000    ?      0.000    ?      ?      0.801    0.014    2
0.455    0.012    0.526    0.455    0.488    0.476    0.899    0.545    3
0.590    0.046    0.512    0.590    0.548    0.510    0.918    0.610    4
0.616    0.066    0.715    0.616    0.662    0.582    0.922    0.781    5
0.891    0.204    0.830    0.891    0.859    0.691    0.941    0.959    6
Weighted Avg.   0.740    0.136    ?      0.740    ?      ?      0.927    0.831
```

Expérience 2

Comme l'expérience 1, sauf que IDfTransform est mis à Faux.

```
=== Summary ===
Correctly Classified Instances      1753      74.0287 %
Incorrectly Classified Instances    615      25.9713 %
Kappa statistic                    0.5898
Mean absolute error                0.0948
Root mean squared error            0.2444
Relative absolute error            43.9558 %
Root relative squared error        74.4643 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
195    0    6   39   31   90 |  a = 1
  1    0    0    1    4    5 |  b = 2
   3    0   30   12    7   14 |  c = 3
  21    0    4  105   14   34 |  d = 4
  67    0   13   27  308   85 |  e = 5
  45    0    4   21   67 1115 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.540    0.068    0.587    0.540    0.563    0.489    0.901    0.644    1
0.000    0.000    ?      0.000    ?      ?      0.801    0.014    2
0.455    0.012    0.526    0.455    0.488    0.476    0.899    0.545    3
0.590    0.046    0.512    0.590    0.548    0.510    0.918    0.610    4
0.616    0.066    0.715    0.616    0.662    0.582    0.922    0.781    5
0.891    0.204    0.830    0.891    0.859    0.691    0.941    0.959    6
Weighted Avg.   0.740    0.136    ?      0.740    ?      ?      0.927    0.831
```

Expérience 3

Comme l'expérience 1, sauf que MultiStopwords n'est pas utilisé.

```

=== Summary ===
Correctly Classified Instances      1753      74.0287 %
Incorrectly Classified Instances    615      25.9713 %
Kappa statistic                    0.5898
Mean absolute error                0.0948
Root mean squared error            0.2444
Relative absolute error            43.9558 %
Root relative squared error        74.4643 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
195   0   6   39  31  90 |  a = 1
  1   0   0   1   4   5 |  b = 2
  3   0  30  12   7  14 |  c = 3
 21   0   4 105  14  34 |  d = 4
 67   0  13  27 308  85 |  e = 5
 45   0   4  21  67 1115 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.540   0.068   0.587    0.540   0.563    0.489   0.901   0.644    1
0.000   0.000   ?      0.000   ?      ?      0.801   0.014    2
0.455   0.012   0.526    0.455   0.488    0.476   0.899   0.545    3
0.590   0.046   0.512    0.590   0.548    0.510   0.918   0.610    4
0.616   0.066   0.715    0.616   0.662    0.582   0.922   0.781    5
0.891   0.204   0.830    0.891   0.859    0.691   0.941   0.959    6
Weighted Avg.   0.740   0.136   ?      0.740   ?      ?      0.927   0.831

```

Expérience 4

Comme l'expérience 1, sauf que SnowballStemmer n'est pas utilisé.

```

=== Summary ===
Correctly Classified Instances      1752      73.9865 %
Incorrectly Classified Instances    616      26.0135 %
Kappa statistic                    0.5943
Mean absolute error                0.0996
Root mean squared error            0.249
Relative absolute error            46.2004 %
Root relative squared error        75.8722 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
202   0  10  28  39  82 |  a = 1
  3   0   0   1   6   1 |  b = 2
  6   0  31   9   8  12 |  c = 3
 26   0   8  92  16  36 |  d = 4
 70   0  12  24 335  59 |  e = 5
 58   0   8  17  77 1092 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.560   0.081   0.553    0.560   0.556    0.476   0.889   0.626    1
0.000   0.000   ?      0.000   ?      ?      0.837   0.018    2
0.470   0.017   0.449    0.470   0.459    0.444   0.897   0.529    3
0.517   0.036   0.538    0.517   0.527    0.490   0.903   0.578    4
0.670   0.078   0.696    0.670   0.683    0.600   0.912   0.763    5
0.872   0.170   0.852    0.872   0.862    0.703   0.935   0.953    6
Weighted Avg.   0.740   0.122   ?      0.740   ?      ?      0.919   0.819

```

Expérience 5

Comme l'expérience 1, sauf que le fichier stoplist-english.txt n'est pas utilisé.

```

=== Summary ===
Correctly Classified Instances      1852      78.2095 %
Incorrectly Classified Instances    516      21.7905 %
Kappa statistic                    0.6685
Mean absolute error                0.0773
Root mean squared error            0.2362
Relative absolute error            35.8389 %
Root relative squared error        71.9649 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
237   0  13  34  48  29 |  a = 1
  3   4   0   2   0   2 |  b = 2
  6   0  35  10   5  10 |  c = 3
 11   0   7 123  19  18 |  d = 4
 56   0  15  34 354  41 |  e = 5
 52   0  12  32  57 1099 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.657   0.064   0.649    0.657   0.653    0.590   0.931   0.720    1
0.364   0.000   1.000    0.364   0.533    0.602   0.905   0.415    2
0.530   0.020   0.427    0.530   0.473    0.459   0.915   0.554    3
0.691   0.051   0.523    0.691   0.596    0.564   0.944   0.710    4
0.708   0.069   0.733    0.708   0.720    0.647   0.928   0.818    5
0.878   0.090   0.917    0.878   0.897    0.787   0.966   0.974    6
Weighted Avg.   0.782   0.076   0.794    0.782   0.786    0.701   0.949   0.868

```

Expérience 6

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés.

```
=== Summary ===
Correctly Classified Instances      1852      78.2095 %
Incorrectly Classified Instances    516      21.7905 %
Kappa statistic                    0.6676
Mean absolute error                 0.0773
Root mean squared error             0.237
Relative absolute error             35.8523 %
Root relative squared error         72.2325 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
238      0      13      33      46      31 |  a = 1
 4       4       0       2       0       1 |  b = 2
 6       0      35      10       4      11 |  c = 3
12       0       8     121      18      19 |  d = 4
58       0      12      32     352      46 |  e = 5
49       0      12      33      56    1102 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.659   0.064   0.649   0.659   0.654   0.591   0.929   0.713   1
0.364   0.000   1.000   0.364   0.533   0.602   0.921   0.421   2
0.530   0.020   0.438   0.530   0.479   0.465   0.921   0.556   3
0.680   0.050   0.524   0.680   0.592   0.559   0.945   0.711   4
0.704   0.066   0.739   0.704   0.721   0.649   0.930   0.818   5
0.880   0.097   0.911   0.880   0.895   0.782   0.965   0.973   6
Weighted Avg.   0.782   0.079   0.793   0.782   0.786   0.699   0.949   0.867
```

Expérience 7

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 1.

```
=== Summary ===
Correctly Classified Instances      1800      76.0135 %
Incorrectly Classified Instances    568      23.9865 %
Kappa statistic                    0.6401
Mean absolute error                 0.0868
Root mean squared error             0.2414
Relative absolute error             40.2593 %
Root relative squared error         73.5633 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
253      4      22      29      31      22 |  a = 1
 5       3       0       2       0       1 |  b = 2
 7       1      35       7       7       9 |  c = 3
26       7      17     102      15      11 |  d = 4
99       2      35      24     309      31 |  e = 5
55       9      20      38      32    1098 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.701   0.096   0.569   0.701   0.628   0.557   0.921   0.674   1
0.273   0.010   0.115   0.273   0.162   0.172   0.945   0.223   2
0.530   0.041   0.271   0.530   0.359   0.355   0.902   0.401   3
0.573   0.046   0.505   0.573   0.537   0.498   0.923   0.588   4
0.618   0.046   0.784   0.618   0.691   0.627   0.927   0.811   5
0.877   0.066   0.937   0.877   0.906   0.809   0.972   0.979   6
Weighted Avg.   0.760   0.064   0.794   0.760   0.772   0.693   0.949   0.848
```

Expérience 8

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 3.

```
=== Summary ===
Correctly Classified Instances      1858      78.4628 %
Incorrectly Classified Instances    510      21.5372 %
Kappa statistic                    0.6703
Mean absolute error                 0.0781
Root mean squared error             0.2363
Relative absolute error             36.2235 %
Root relative squared error         72.0133 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
234      0      12      26      49      40 |  a = 1
 1       3       0       2       1       4 |  b = 2
 4       0      36      12       6       8 |  c = 3
18       0      10     117       9      24 |  d = 4
53       1      16      29     362      39 |  e = 5
48       0       8      32      58    1106 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.648   0.062   0.654   0.648   0.651   0.588   0.928   0.726   1
0.273   0.000   0.750   0.273   0.400   0.451   0.881   0.369   2
0.545   0.020   0.439   0.545   0.486   0.473   0.909   0.567   3
0.657   0.046   0.537   0.657   0.591   0.557   0.936   0.695   4
0.724   0.066   0.746   0.724   0.735   0.666   0.922   0.813   5
0.883   0.103   0.906   0.883   0.894   0.779   0.959   0.966   6
Weighted Avg.   0.785   0.082   0.792   0.785   0.787   0.700   0.943   0.863
```

Expérience 9

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 5.

```
=== Summary ===
Correctly Classified Instances      1878      79.3074 %
Incorrectly Classified Instances    490      20.6926 %
Kappa statistic                    0.681
Mean absolute error                0.0735
Root mean squared error            0.2327
Relative absolute error            34.1132 %
Root relative squared error        70.9054 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.712    0.068    0.652     0.712    0.681     0.621    0.934     0.753     1
0.000    0.000    ?         0.000    ?         ?         0.773     0.188     2
0.515    0.020    0.425     0.515    0.466     0.451    0.927     0.568     3
0.618    0.032    0.608     0.618    0.613     0.581    0.935     0.707     4
0.718    0.060    0.762     0.718    0.739     0.673    0.929     0.828     5
0.893    0.111    0.900     0.893    0.897     0.782    0.961     0.969     6
Weighted Avg.    0.793    0.085    ?         0.793    ?         ?         0.946     0.872
```

Analyse

IDFTransform n'a pas d'impact et, quand stoplist-english.txt est utilisé, MultiStopwords n'a aussi pas d'impact.

Quand SnowballStemmer n'est pas utilisé, le taux d'instances correctement classifiées diminue de 0.0422%. Pour environ la moitié des classes, la précision augmente légèrement tandis que, pour l'autre moitié, elle diminue légèrement. C'est la même chose pour le rappel. Alors, le taux d'instances correctement classifiées pour chaque classe varie un peu, mais, globalement, le taux diminue faiblement.

Ne pas utiliser le fichier stoplist-english.txt augmente le taux d'instances correctement classifiées de 4.18%. Il permet, en outre, de classer correctement quelques instances avec le deuxième sens tandis qu'avec le stoplist-english.txt, aucune instance de cette catégorie n'était identifiée. De plus, si stoplist-english.txt et MultiStopwords ne sont pas utilisés, les classifications changent très légèrement, mais le taux reste exactement le même.

Diminuer la taille de la fenêtre à 1 diminue le taux d'instances correctement classifiées de 2.196%. Augmenter la taille à 3 augmente le taux de 0.2533% et Augmenter la taille à 5 augmente le taux de 1.098%. Une fenêtre plus grande donne, donc, de meilleurs résultats.

La meilleure performance obtenue est, alors, un taux d'instances correctement classifiées de 79.3074%.

Arbre de décision (J48)

Expérience 1

La taille de la fenêtre de contexte est de 2 mots et le fichier stoplist-english.txt est utilisé pour enlever les mots outils. Dans le preprocess de Weka, IDFTTransform, TFTransform et lowerCaseToken sont à True. De plus, SnowballStemmer et MultiStopwords sont utilisés.

```
=== Summary ===
Correctly Classified Instances      1888      79.7297 %
Incorrectly Classified Instances    480      20.2703 %
Kappa statistic                    0.6886
Mean absolute error                 0.077
Root mean squared error             0.2167
Relative absolute error             35.7293 %
Root relative squared error         66.025 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a    b    c    d    e    f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6
      a    b    c    d    e    f
a      6     0     0     1     4     0
b      21    0    30     3    11     1
c      57    0     3    79    32     7
d      69    0     4     7   399    21
e      62     1     3     6   42  1138
f

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.670   0.107   0.530   0.670   0.592   0.513   0.900   0.647    1
0.000   0.000   0.000   0.000   0.000  -0.001   0.717   0.023    2
0.455   0.006   0.682   0.455   0.545   0.547   0.871   0.484    3
0.444   0.015   0.712   0.444   0.547   0.535   0.874   0.575    4
0.798   0.090   0.702   0.798   0.747   0.676   0.936   0.833    5
0.909   0.044   0.959   0.909   0.933   0.864   0.968   0.974    6
Weighted Avg.   0.797   0.060   0.808   0.797   0.798   0.733   0.940   0.846
```

Expérience 2

Comme l'expérience 1, sauf que IDFTTransform est mis à Faux.

```
=== Summary ===
Correctly Classified Instances      1888      79.7297 %
Incorrectly Classified Instances    480      20.2703 %
Kappa statistic                    0.6886
Mean absolute error                 0.077
Root mean squared error             0.2167
Relative absolute error             35.7293 %
Root relative squared error         66.025 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a    b    c    d    e    f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6
      a    b    c    d    e    f
a      6     0     0     1     4     0
b      21    0    30     3    11     1
c      57    0     3    79    32     7
d      69    0     4     7   399    21
e      62     1     3     6   42  1138
f

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.670   0.107   0.530   0.670   0.592   0.513   0.900   0.647    1
0.000   0.000   0.000   0.000   0.000  -0.001   0.717   0.023    2
0.455   0.006   0.682   0.455   0.545   0.547   0.871   0.484    3
0.444   0.015   0.712   0.444   0.547   0.535   0.874   0.575    4
0.798   0.090   0.702   0.798   0.747   0.676   0.936   0.833    5
0.909   0.044   0.959   0.909   0.933   0.864   0.968   0.974    6
Weighted Avg.   0.797   0.060   0.808   0.797   0.798   0.733   0.940   0.846
```

Expérience 3

Comme l'expérience 1, sauf que SnowballStemmer n'est pas utilisé.

```

=== Summary ===
Correctly Classified Instances      1852      78.2095 %
Incorrectly Classified Instances    516      21.7905 %
Kappa statistic                    0.6652
Mean absolute error                0.0843
Root mean squared error            0.2256
Relative absolute error            39.0886 %
Root relative squared error        68.7553 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a  b  c  d  e  f  <-- classified as
225  0  5  25  80  26 |  a = 1
 4  0  0  1  6  0 |  b = 2
16  0  28  5  15  2 |  c = 3
49  0  7  73  34  15 |  d = 4
58  0  3  17  400  22 |  e = 5
62  0  5  13  46  1126 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.623  0.094  0.543  0.623  0.581  0.501  0.889  0.631  1
0.000  0.000  ?  0.000  ?  ?  0.834  0.021  2
0.424  0.009  0.583  0.424  0.491  0.485  0.853  0.405  3
0.410  0.028  0.545  0.410  0.468  0.436  0.829  0.494  4
0.800  0.097  0.688  0.800  0.740  0.667  0.927  0.808  5
0.899  0.058  0.945  0.899  0.922  0.840  0.964  0.969  6
Weighted Avg.  0.782  0.068  ?  0.782  ?  ?  0.931  0.828

```

Expérience 4

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés.

```

=== Summary ===
Correctly Classified Instances      1989      83.9949 %
Incorrectly Classified Instances    379      16.0051 %
Kappa statistic                    0.7525
Mean absolute error                0.0635
Root mean squared error            0.2024
Relative absolute error            29.4453 %
Root relative squared error        61.6657 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a  b  c  d  e  f  <-- classified as
278  1  4  16  39  23 |  a = 1
 4  2  0  2  1  2 |  b = 2
17  0  29  7  8  5 |  c = 3
29  0  5  111  15  18 |  d = 4
64  1  5  13  398  19 |  e = 5
40  0  4  12  25  1171 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.770  0.077  0.644  0.770  0.701  0.645  0.912  0.704  1
0.182  0.001  0.500  0.182  0.267  0.300  0.759  0.171  2
0.439  0.008  0.617  0.439  0.513  0.509  0.842  0.488  3
0.624  0.023  0.689  0.624  0.655  0.629  0.896  0.635  4
0.796  0.047  0.819  0.796  0.807  0.757  0.940  0.852  5
0.935  0.060  0.946  0.935  0.941  0.875  0.975  0.978  6
Weighted Avg.  0.840  0.055  0.842  0.840  0.839  0.784  0.948  0.867

```

Expérience 5

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 1.

```

=== Summary ===
Correctly Classified Instances      1997      84.3328 %
Incorrectly Classified Instances    371      15.6672 %
Kappa statistic                    0.7586
Mean absolute error                0.0692
Root mean squared error            0.1991
Relative absolute error            32.1128 %
Root relative squared error        60.6593 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a  b  c  d  e  f  <-- classified as
283  0  1  24  36  17 |  a = 1
 4  2  0  4  0  1 |  b = 2
15  0  33  6  5  7 |  c = 3
38  1  1  107  25  6 |  d = 4
76  0  6  14  391  13 |  e = 5
42  0  0  11  18  1181 |  f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.784  0.087  0.618  0.784  0.691  0.634  0.915  0.689  1
0.182  0.000  0.667  0.182  0.286  0.347  0.784  0.175  2
0.500  0.003  0.805  0.500  0.617  0.627  0.874  0.488  3
0.601  0.027  0.645  0.601  0.622  0.593  0.917  0.628  4
0.782  0.045  0.823  0.782  0.802  0.751  0.950  0.866  5
0.943  0.039  0.964  0.943  0.954  0.903  0.980  0.979  6
Weighted Avg.  0.843  0.046  0.852  0.843  0.844  0.796  0.955  0.867

```


Expérience 6

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 3.

```
=== Summary ===
Correctly Classified Instances      1970      83.1926 %
Incorrectly Classified Instances    398       16.8074 %
Kappa statistic                    0.7409
Mean absolute error                0.0638
Root mean squared error            0.2091
Relative absolute error            29.5942 %
Root relative squared error        63.716 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.762    0.083    0.624     0.762    0.686     0.627    0.909     0.656     1
0.091    0.000    0.500     0.091    0.154     0.212    0.619     0.123     2
0.439    0.006    0.690     0.439    0.537     0.541    0.822     0.447     3
0.612    0.024    0.673     0.612    0.641     0.614    0.871     0.615     4
0.778    0.059    0.778     0.778    0.778     0.719    0.925     0.817     5
0.932    0.048    0.956     0.932    0.944     0.883    0.971     0.975     6
Weighted Avg.    0.832    0.053    0.837     0.832    0.832     0.776    0.938     0.847

=== Detailed Accuracy By Class ===
```

Expérience 7

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 5.

```
=== Summary ===
Correctly Classified Instances      1968      83.1081 %
Incorrectly Classified Instances    400       16.8919 %
Kappa statistic                    0.7416
Mean absolute error                0.062
Root mean squared error            0.2125
Relative absolute error            28.7493 %
Root relative squared error        64.7652 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.789    0.086    0.624     0.789    0.697     0.641    0.905     0.633     1
0.091    0.003    0.125     0.091    0.105     0.103    0.614     0.112     2
0.545    0.010    0.621     0.545    0.581     0.571    0.883     0.498     3
0.624    0.034    0.600     0.624    0.612     0.579    0.877     0.615     4
0.746    0.041    0.829     0.746    0.785     0.733    0.909     0.804     5
0.928    0.043    0.960     0.928    0.944     0.884    0.973     0.976     6
Weighted Avg.    0.831    0.047    0.841     0.831    0.834     0.780    0.938     0.843

=== Detailed Accuracy By Class ===
```

Analyse

IDFTransform n'a pas d'impact. Quand SnowballStemmer n'est pas utilisé, le taux d'instances correctement classifiées diminue de 1.5202%.

Ne pas utiliser stoplist-english.txt et MultiStopwords augmente le taux de 4.2652%. De plus, diminuer la taille de la fenêtre à 1 permet d'augmenter le taux de 0.3379%.

Augmenter la fenêtre à 3 mots descend le taux de 0.8023% et augmenter la fenêtre à 5 mots fait descendre le taux de 0.8868%. Donc, l'arbre de décision performe le mieux avec une fenêtre de 1 mot et performe de moins en moins bien en augmentant la taille de la fenêtre.

La meilleure performance obtenue est, alors, un taux d'instances correctement classifiées de 84.3328%.

SVM (SMO)

Expérience 1

La taille de la fenêtre de contexte est de 2 mots et le fichier stoplist-english.txt est utilisé pour enlever les mots outils. Dans le preprocess de Weka, IDFFTransform, TFTransform et lowerCaseToken sont à True. De plus, SnowballStemmer et MultiStopwords sont utilisés.

```
=== Summary ===                               === Confusion Matrix ===

Correctly Classified Instances      2088      88.1757 %
Incorrectly Classified Instances    280      11.8243 %
Kappa statistic                    0.8173
Mean absolute error                 0.2271
Root mean squared error             0.318
Relative absolute error             105.3492 %
Root relative squared error         96.8917 %
Total Number of Instances          2368

a    b    c    d    e    f  <-- classified as
296  0    2    20   28   15 |  a = 1
 8    0    0    1    2    0 |  b = 2
16    0   34    6    8    2 |  c = 3
44    0    6   111   15    2 |  d = 4
30    0    1    13   448    8 |  e = 5
28    0    0    4    21  1199 |  f = 6

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.820    0.063    0.701    0.820    0.756    0.711    0.894    0.624    1
      0.000    0.000    ?        0.000    ?        ?        0.489    0.005    2
      0.515    0.004    0.791    0.515    0.624    0.630    0.830    0.490    3
      0.624    0.020    0.716    0.624    0.667    0.643    0.892    0.548    4
      0.896    0.040    0.858    0.896    0.877    0.843    0.943    0.804    5
      0.958    0.024    0.978    0.958    0.968    0.933    0.977    0.968    6
Weighted Avg.  0.882    0.032    ?        0.882    ?        ?        0.945    0.832
```

Expérience 2

Comme l'expérience 1, sauf que TFTransform est mis à Faux.

```
=== Summary ===                               === Confusion Matrix ===

Correctly Classified Instances      2088      88.1757 %
Incorrectly Classified Instances    280      11.8243 %
Kappa statistic                    0.8173
Mean absolute error                 0.2271
Root mean squared error             0.318
Relative absolute error             105.3492 %
Root relative squared error         96.8917 %
Total Number of Instances          2368

a    b    c    d    e    f  <-- classified as
296  0    2    20   28   15 |  a = 1
 8    0    0    1    2    0 |  b = 2
16    0   34    6    8    2 |  c = 3
44    0    6   111   15    2 |  d = 4
30    0    1    13   448    8 |  e = 5
28    0    0    4    21  1199 |  f = 6

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.820    0.063    0.701    0.820    0.756    0.711    0.894    0.624    1
      0.000    0.000    ?        0.000    ?        ?        0.489    0.005    2
      0.515    0.004    0.791    0.515    0.624    0.630    0.830    0.490    3
      0.624    0.020    0.716    0.624    0.667    0.643    0.892    0.548    4
      0.896    0.040    0.858    0.896    0.877    0.843    0.943    0.804    5
      0.958    0.024    0.978    0.958    0.968    0.933    0.977    0.968    6
Weighted Avg.  0.882    0.032    ?        0.882    ?        ?        0.945    0.832
```

Expérience 3

Comme l'expérience 1, sauf que SnowballStemmer n'est pas utilisé.

```

=== Summary ===
Correctly Classified Instances      2062      87.0777 %
Incorrectly Classified Instances    306      12.9223 %
Kappa statistic                    0.7995
Mean absolute error                0.2277
Root mean squared error            0.319
Relative absolute error            105.6452 %
Root relative squared error        97.203 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

```

```

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.801   0.065   0.688   0.801   0.740   0.692   0.886   0.597   1
0.000   0.000   ?       0.000   ?       ?       0.490   0.005   2
0.500   0.004   0.767   0.500   0.606   0.611   0.811   0.464   3
0.579   0.021   0.696   0.579   0.632   0.608   0.860   0.503   4
0.882   0.040   0.855   0.882   0.868   0.832   0.933   0.793   5
0.955   0.040   0.964   0.955   0.959   0.914   0.966   0.952   6
Weighted Avg.   0.871   0.041   ?       0.871   ?       ?       0.932   0.813

```

Expérience 4

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés.

```

=== Summary ===
Correctly Classified Instances      2107      88.978 %
Incorrectly Classified Instances    261      11.022 %
Kappa statistic                    0.8287
Mean absolute error                0.2267
Root mean squared error            0.317
Relative absolute error            105.1446 %
Root relative squared error        96.5939 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

```

```

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.798   0.039   0.785   0.798   0.791   0.753   0.910   0.695   1
0.273   0.001   0.600   0.273   0.375   0.403   0.838   0.213   2
0.561   0.007   0.685   0.561   0.617   0.610   0.827   0.415   3
0.702   0.022   0.718   0.702   0.710   0.687   0.906   0.578   4
0.876   0.037   0.864   0.876   0.870   0.835   0.945   0.804   5
0.971   0.040   0.964   0.971   0.968   0.931   0.970   0.957   6
Weighted Avg.   0.890   0.037   0.888   0.890   0.888   0.854   0.946   0.838

```

Expérience 5

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 1.

```

=== Summary ===
Correctly Classified Instances      2066      87.2466 %
Incorrectly Classified Instances    302      12.7534 %
Kappa statistic                    0.8041
Mean absolute error                0.2273
Root mean squared error            0.318
Relative absolute error            105.4188 %
Root relative squared error        96.9064 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

```

```

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.784   0.050   0.737   0.784   0.760   0.715   0.890   0.617   1
0.000   0.000   ?       0.000   ?       ?       0.753   0.090   2
0.530   0.005   0.745   0.530   0.619   0.620   0.772   0.408   3
0.719   0.037   0.615   0.719   0.663   0.636   0.909   0.513   4
0.872   0.042   0.848   0.872   0.860   0.822   0.945   0.799   5
0.946   0.028   0.974   0.946   0.960   0.917   0.974   0.964   6
Weighted Avg.   0.872   0.034   ?       0.872   ?       ?       0.944   0.823

```

Expérience 6

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 3.

```
=== Summary ===
Correctly Classified Instances      2122      89.6115 %
Incorrectly Classified Instances    246       10.3885 %
Kappa statistic                    0.8384
Mean absolute error                 0.2264
Root mean squared error             0.3166
Relative absolute error             105.027 %
Root relative squared error         96.4772 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6
297  0   2  15  33  14 |
  4   2   0   1   1   3 |
  9   0  36  10   8   3 |
 16   0   5 132  13  12 |
 34   0   0  11 441  14 |
 19   0   0   7  12 1214 |

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.823    0.041    0.784    0.823    0.803    0.767    0.922    0.703    1
0.182    0.000    1.000    0.182    0.308    0.426    0.714    0.334    2
0.545    0.003    0.837    0.545    0.661    0.669    0.845    0.502    3
0.742    0.020    0.750    0.742    0.746    0.725    0.909    0.617    4
0.882    0.036    0.868    0.882    0.875    0.841    0.942    0.809    5
0.970    0.041    0.963    0.970    0.967    0.929    0.972    0.957    6
Weighted Avg.    0.896    0.037    0.897    0.896    0.894    0.861    0.948    0.846
```

Expérience 7

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés et la taille de la fenêtre est 5.

```
=== Summary ===
Correctly Classified Instances      2115      89.3159 %
Incorrectly Classified Instances    253       10.6841 %
Kappa statistic                    0.8339
Mean absolute error                 0.2264
Root mean squared error             0.3167
Relative absolute error             105.0096 %
Root relative squared error         96.4934 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a   b   c   d   e   f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6
294  1   0  18  36  12 |
  6   0   0   1   3   1 |
  8   0  38   9  10   1 |
 24   1   5 124  11  13 |
 33   0   2   9 443  13 |
 18   0   1   4  13 1216 |

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.814    0.044    0.768    0.814    0.790    0.752    0.922    0.686    1
0.000    0.001    0.000    0.000    0.000   -0.002    0.749    0.096    2
0.576    0.003    0.826    0.576    0.679    0.683    0.858    0.553    3
0.697    0.019    0.752    0.697    0.723    0.702    0.902    0.596    4
0.886    0.039    0.859    0.886    0.872    0.837    0.944    0.803    5
0.971    0.036    0.968    0.971    0.970    0.936    0.975    0.962    6
Weighted Avg.    0.893    0.035    0.890    0.893    0.891    0.858    0.950    0.843
```

Analyse

TFTransform n'a pas d'impact. Quand SnowballStemmer n'est pas utilisé, le taux d'instances correctement classifiées diminue de 1.098%.

Ne pas utiliser le fichier stoplist-english.txt et MultiStopwords augmente le taux d'instances correctement classifiées de 0.8023%. Il permet, en outre, de classer correctement quelques instances avec le deuxième sens tandis que sans les stopwords, aucune instance de cette catégorie n'était identifiée.

Diminuer la taille de la fenêtre de contexte à 1 mots diminue le taux de 1.7314%. Augmenter la taille à 3 mots augmente le taux de 0.6335% et augmenter la fenêtre à 5 mots donne une augmentation du taux de 0.3379%. Donc, le SVM a sa meilleure performance avec une fenêtre de 3 mots.

La meilleure performance obtenue est, alors, un taux d'instances correctement classifiées de 89.6115%.

MultiLayerPerceptron

Expérience 1

La taille de la fenêtre de contexte est de 2 mots et le fichier stoplist-english.txt est utilisé pour enlever les mots outils. Dans le preprocess de Weka, IDFTTransform, TFTransform et lowerCaseToken sont à True. De plus, SnowballStemmer et MultiStopwords sont utilisés. Pour ce qui est des caractéristiques particulières au MultiLayerPerceptron, il y a 5 neurones cachés et reset est à False.

```
=== Summary ===
Correctly Classified Instances      1252      52.8716 %
Incorrectly Classified Instances    1116      47.1284 %
Kappa statistic                     0
Mean absolute error                 0.2154
Root mean squared error             0.3301
Relative absolute error             99.931 %
Root relative squared error         100.5839 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.000    0.000    ?          0.000    ?          ?        0.499    0.147    1
0.000    0.000    ?          0.000    ?          ?        0.448    0.005    2
0.000    0.000    ?          0.000    ?          ?        0.483    0.027    3
0.000    0.000    ?          0.000    ?          ?        0.485    0.072    4
0.000    0.000    ?          0.000    ?          ?        0.510    0.248    5
1.000    1.000    0.529      1.000    0.692      ?        0.505    0.512    6
Weighted Avg.  0.529    0.529    ?          0.529    ?          ?        0.503    0.351
```

Expérience 2

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés, la taille de la fenêtre est 5 et reset est mis à True. (avec 10 et 30 neurones cachés, ça donne le même résultat)

```
=== Summary ===
Correctly Classified Instances      1252      52.8716 %
Incorrectly Classified Instances    1116      47.1284 %
Kappa statistic                     0
Mean absolute error                 0.2168
Root mean squared error             0.3298
Relative absolute error             100.5836 %
Root relative squared error         100.4924 %
Total Number of Instances          2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
a = 1
b = 2
c = 3
d = 4
e = 5
f = 6

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.000    0.000    ?          0.000    ?          ?        0.509    0.156    1
0.000    0.000    ?          0.000    ?          ?        0.460    0.005    2
0.000    0.000    ?          0.000    ?          ?        0.494    0.028    3
0.000    0.000    ?          0.000    ?          ?        0.498    0.085    4
0.000    0.000    ?          0.000    ?          ?        0.512    0.211    5
1.000    1.000    0.529      1.000    0.692      ?        0.510    0.570    6
Weighted Avg.  0.529    0.529    ?          0.529    ?          ?        0.509    0.377
```

Expérience 3

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés, il y a 30 neurones cachés, reset est mis à False et le learning rate est à 0.1 au lieu de 0.3.

```

=== Summary ===
Correctly Classified Instances      1279      54.0118 %
Incorrectly Classified Instances    1089      45.9882 %
Kappa statistic                    0.0575
Mean absolute error                0.2102
Root mean squared error            0.3251
Relative absolute error             97.521 %
Root relative squared error        99.0753 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
0      0      0      0      34     327 | a = 1
0      0      0      0      1      10 | b = 2
0      0      0      0      5      61 | c = 3
0      0      0      0     17     161 | d = 4
0      0      0      0     47     453 | e = 5
0      0      0      0     20    1232 | f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.000  0.000  ?      0.000  ?      ?      0.572  0.192  1
0.000  0.000  ?      0.000  ?      ?      0.452  0.005  2
0.000  0.000  ?      0.000  ?      ?      0.547  0.034  3
0.000  0.000  ?      0.000  ?      ?      0.558  0.101  4
0.094  0.041  0.379  0.094  0.151  0.097  0.582  0.298  5
0.984  0.907  0.549  0.984  0.705  0.173  0.612  0.668  6
Weighted Avg.  0.540  0.488  ?      0.540  ?      ?      0.593  0.454

```

Expérience 4

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés, il y a 10 neurones cachés, reset est mis à False, TrainingTime est à 100 au lieu de 500 et le learning rate est à 0.01 au lieu de 0.3.

```

=== Summary ===
Correctly Classified Instances      1951      82.3902 %
Incorrectly Classified Instances    417      17.6098 %
Kappa statistic                    0.7231
Mean absolute error                0.106
Root mean squared error            0.2083
Relative absolute error             49.1926 %
Root relative squared error        63.479 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
254      0      0     30     51     26 | a = 1
3      0      0      5      1      2 | b = 2
11      0      0     30     12     13 | c = 3
39      0      0     93     22     24 | d = 4
51      0      0     23    401     25 | e = 5
26      0      0     12     11    1203 | f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.704  0.065  0.661  0.704  0.682  0.623  0.933  0.744  1
0.000  0.000  ?      0.000  ?      ?      0.705  0.026  2
0.000  0.000  ?      0.000  ?      ?      0.804  0.097  3
0.522  0.046  0.482  0.522  0.501  0.459  0.907  0.489  4
0.802  0.052  0.805  0.802  0.804  0.751  0.962  0.887  5
0.961  0.081  0.930  0.961  0.945  0.883  0.985  0.988  6
Weighted Avg.  0.824  0.067  ?      0.824  ?      ?      0.960  0.862

```

Expérience 5

Comme l'expérience 1, sauf que le fichier stoplist-english.txt et le MultiStopwords ne sont pas utilisés, il y a 30 neurones cachés, reset est mis à False, TrainingTime est à 100 au lieu de 500 et le learning rate est à 0.01 au lieu de 0.3.

```

=== Summary ===
Correctly Classified Instances      1981      83.6571 %
Incorrectly Classified Instances    387      16.3429 %
Kappa statistic                    0.7424
Mean absolute error                0.0835
Root mean squared error            0.1943
Relative absolute error             38.7481 %
Root relative squared error        59.211 %
Total Number of Instances         2368

=== Confusion Matrix ===
      a      b      c      d      e      f  <-- classified as
258      0      0     22     52     29 | a = 1
3      0      0      5      2      1 | b = 2
12      0      0     21     18     15 | c = 3
20      0      0    106     29     23 | d = 4
46      0      0     19    413     22 | e = 5
21      0      0      6     21    1204 | f = 6

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.715  0.051  0.717  0.715  0.716  0.665  0.946  0.788  1
0.000  0.000  ?      0.000  ?      ?      0.833  0.018  2
0.000  0.000  ?      0.000  ?      ?      0.879  0.261  3
0.596  0.033  0.592  0.596  0.594  0.561  0.941  0.685  4
0.826  0.065  0.772  0.826  0.798  0.742  0.968  0.904  5
0.962  0.081  0.930  0.962  0.946  0.883  0.988  0.991  6
Weighted Avg.  0.837  0.067  ?      0.837  ?      ?      0.970  0.894

```

Analyse

Dans les premières expériences, le MultiLayerPerceptron n'a pas très bien performé. L'utilisation de stopwords, le nombre de neurones cachés et l'utilisation de reset n'ont pas eu l'air d'avoir d'impact. Il a eu tendance à seulement tout classifier les instances dans la sixième catégorie. Cela pourrait être causé par le déséquilibre dans la représentation des classes dans les instances (53% des instances sont dans la sixième catégorie).

Diminuer le learning rate à 0.1 a semblé aider le MultiLayerPerceptron, mais il a seulement classifié toutes les instances dans 2 des 6 catégories (qui représente 53% et 21% de toutes les instances) au lieu de toutes les classifiées dans une seule catégorie. Diminuer le learning rate à 0.01 permet de classifier les instances dans 4 des 6 catégories (en ignorant les catégories représentant 1% et 3% des instances). De plus, avec ce learning rate, augmenter le nombre de neurones cachés de 10 à 30 a augmenté le taux d'instances correctement classifiées de 1.2669%.

La meilleure performance obtenue est, alors, un taux d'instances correctement classifiées de 83.6571%.

Comparaison entre algorithmes

Dans tous les cas, IDfTransform et TfidfTransform n'ont pas d'impact. De plus, utiliser stoplist-english.txt et MultiStopwords donne étonnamment de moins bons résultats. Quand les stopwords ne sont pas enlevés, le taux d'instances correctement classifiées augmente de 4.18% pour Naives Bayes, 4.2652% pour l'arbre de décision et seulement 0.8023% pour le SVM. SnowballStemmer permet d'améliorer légèrement les résultats (augmente le taux de 0.0422% pour Naives Bayes, 1.5202% pour l'arbre et 1.098% pour le SVM).

L'arbre de décision performe mieux que Naive Bayes en obtenant un taux d'instances correctement classifiées plus grand de 5.0254% (84.3328% pour l'arbre contre 79.3074% pour Naives Bayes). L'arbre performe le mieux avec une fenêtre de 1 mot et performe de moins en moins bien en grandissant la fenêtre tandis que c'est exactement le contraire pour Naive Bayes.

Le MultiLayerPerceptron a mieux performé (avec un taux de 83.6571%) que Naives Bayes, mais légèrement moins que l'arbre de décision qui a eu un taux plus grand de 0.6757%. De plus, il n'a pas réussi à identifier une seule instance appartenant à la troisième catégorie tandis que l'arbre de décision a correctement identifié 33 des 66 instances dans cette catégorie et Naives Bayes 34 des 66.

SVM est la méthode qui a obtenu les meilleurs résultats. Le SVM performe mieux que l'arbre de décision en obtenant un taux d'instances correctement classifiées plus grand de 5.2787% (89.6115% pour le SVM contre 84.3328% pour l'arbre). Contrairement à l'arbre et à Naives Bayes, le SVM a performé le mieux avec une fenêtre de 3 mots.