CSI4107
Assignment 2

Junhan Liu 7228243
Yinggin Guan 7285769

Introduction

Our choice of programming language to extract features from the twitter collection is Python. Our script to generate the according data representation that led to the best result for all three classifiers is located in folder ‘python script’. Generally, the script is able to process each twitter from the collection to get the according features including: positive sentiment score, negative sentiment score, neutral score, compound score, positive/negative words count, contains question mark or not, polarity score of a twitter. Other functionalities for part 1, which uses bag of words representation, includes removing stop words, rare words, emojis, hyperlinks, and all special characters. Detailed explanation of what features are chosen for an experiment is included in each section below. Works were distributed equally. Me and my partner were both programming the script and exploring existing methods to improve our experiment results.

How To Run

Location of script: ‘python script’ folder
Script to run: generate.py

Before you run, you should have python 3.6 installed on you PC. Please do not use Python idle to run the script, under some platforms, it might give you python building error and might result in not able to run script properly.
To be able to get the best experiment result, we applied an external python library called NLTK. The API reference is listed in a REMDME.md file in the according experiment folder. Thus, the necessary external packages might be downloaded before running. Packages you might need to have are: nltk, SentimentIntensityAnalyzer, tweepy, and TextBlob. Once you have downloaded Python, you can use >pip install ‘whatever you need’ to install the missing packages.

Part 1

For part 1, “semeval_twitter_data.arff” from the original package is used for weka preprocessing. StringtoWordVector filter and Remove duplicate is used to make the string to be extracted to numeric type structure for better performance of classification. As showed in Figure 1.1 to 1.2, there are 2369 attributes with unnecessary features (attributes). After that, the 3 classifiers (SVM, NB and DT) is used and the results are showed in Figure 1.3 to 1.5.

Figure 1.1 StringToWordVector

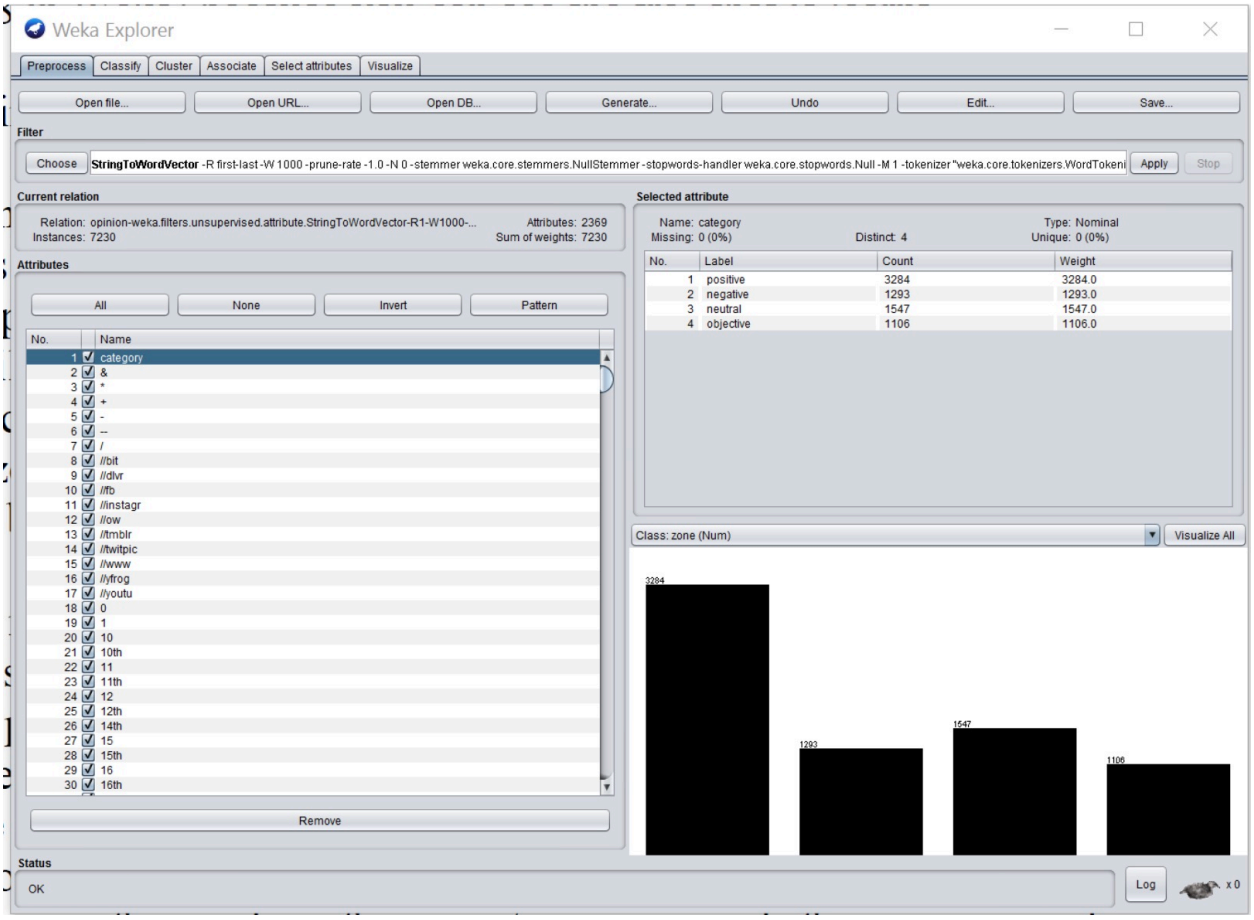
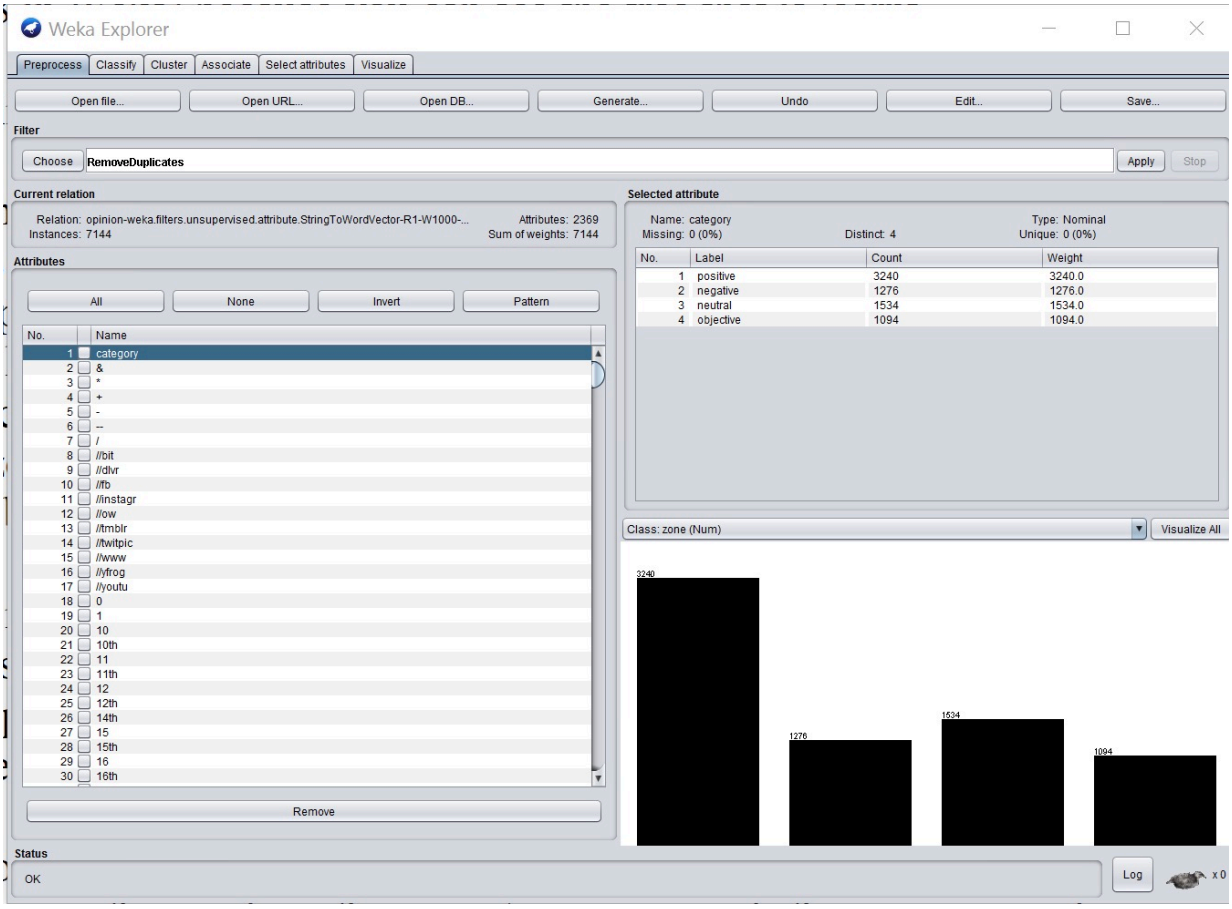


Figure 1.2 RemoveDuplicate



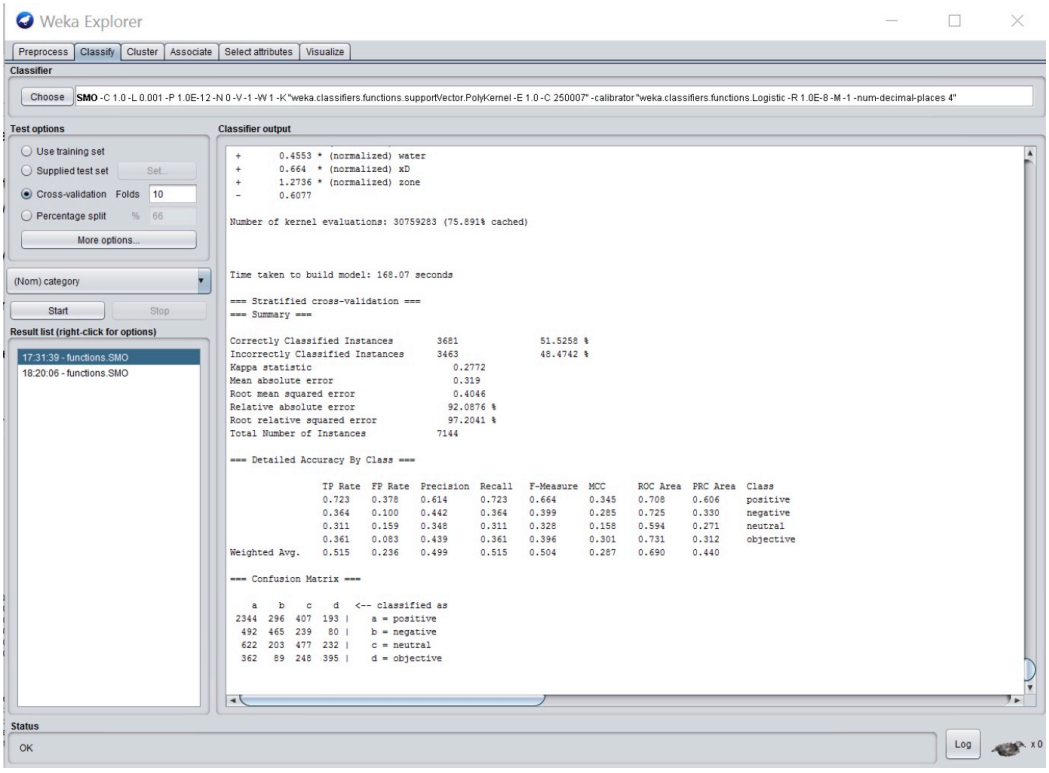


Figure 1.3 SMO

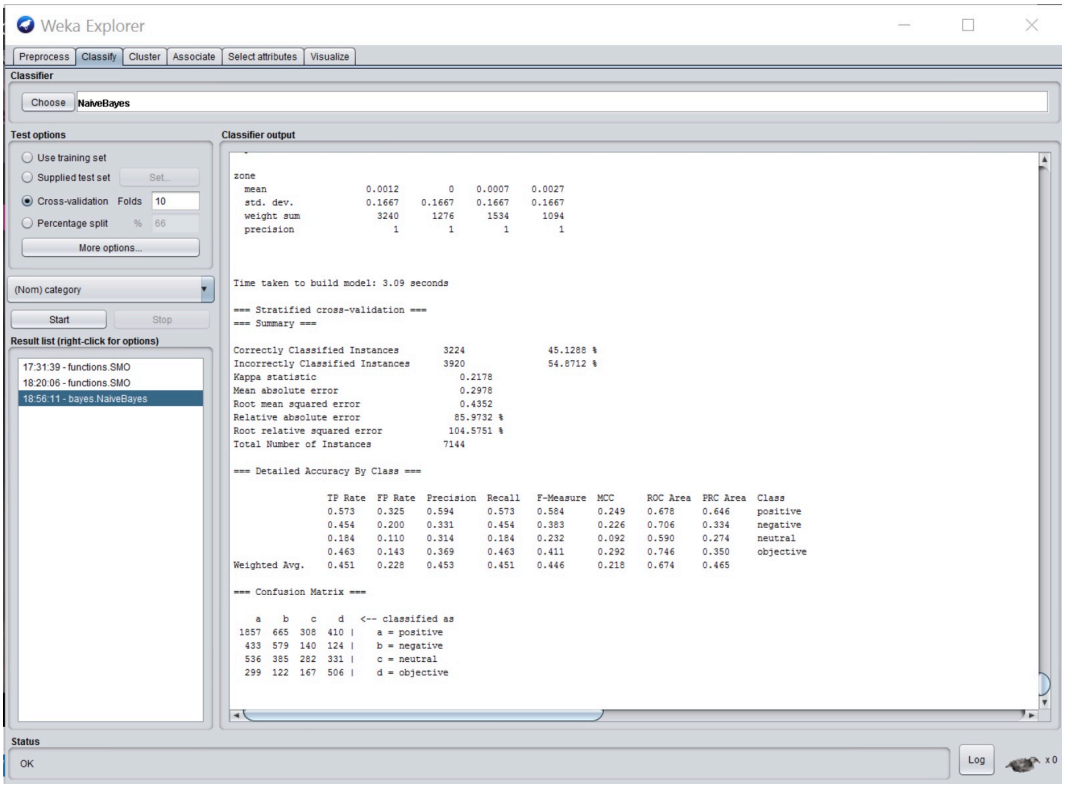
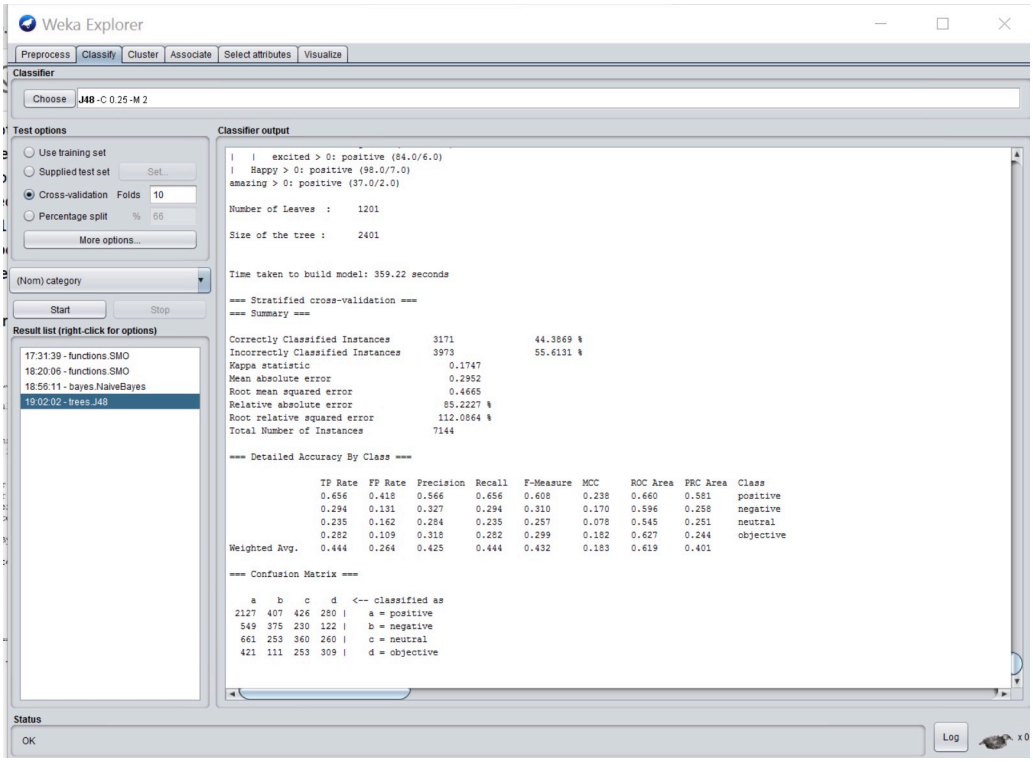


Figure 1.4 NaiveBay



In the this part, we did not add any features to train classifiers. Instead, we simply used bag of words representation. That is we used all words in the message set as attributes to train classifiers including SVM, J48, and Naive Bayes.

- **First Experiment**

Input data file for Weka is the provided twitter file in .arff format: 'semeval_twitter_data.arff'. We kept all words originally in the file, and used StringToWordVector and RemoveDuplicate in Weka to extract words. The overall processing time was longer than expected since dimension of the vector space is high.

For SMO, we have 51.5258% accuracy.
For J48, we have 44.3869% accuracy.
For NB, we have 45.1288% accuracy.

Well, the accuracy of SMO is not so bad. The result is acceptable if not considering the processing time.

- **Second Experiment**

We decided to reduce the dimension of the vector space by using our Python script 'gengerate.py' to eliminate stop words, rare words, punctuation, emoji, and all special characters. The stop word list is located in folder 'resource'.

The outcome was not improved nor declined very significantly. The accuracy was about the same and the processing time was still long.

Part 2

To improve the classification result (accuracy), we did 5 different experiments, which each of them has a distinct set of features for the ML algorithms to train the classifiers. The result was getting better and better and the processing time was reduced significantly.

- **Frist Experiment**

Resource used: Positive word list (Positive.csv), Negative word list (Negative.csv)
Python Script: generate.py
Output folder is 'E1_pos_neg_result'.

Our script counts the number of positive&negative words in each twitter message, then writes data to a new .arff file, which is included in the output folder. The detailed prediction result of three classifiers is included in the output folder.

NB

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3180               43.9834 %
Incorrectly Classified Instances      4050               56.0166 %
Kappa statistic                          0.1099
Mean absolute error                      0.3326
Root mean squared error                  0.4151
Relative absolute error                 96.042  %
Root relative squared error             99.7529 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.759    0.563    0.529      0.759    0.623      0.205    0.634     0.586     positive
                 0.000    0.000    ?          0.000    ?          ?        0.527     0.187     negative
                 0.444    0.321    0.273      0.444    0.338      0.106    0.572     0.248     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.607     0.192     objective
Weighted Avg.    0.440    0.325    ?          0.440    ?          ?        0.597     0.382

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2493    0  791    0 |    a = positive
  790    0  503    0 |    b = negative
  860    0  687    0 |    c = neutral
  573    0  533    0 |    d = objective
```

J48

```
Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3284               45.4219 %
Incorrectly Classified Instances      3946               54.5781 %
Kappa statistic                          0
Mean absolute error                      0.3359
Root mean squared error                  0.4279
Relative absolute error                 96.9977 %
Root relative squared error            102.8399 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.454      1.000    0.625      ?        0.500     0.454     positive
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.179     negative
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.214     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.153     objective
Weighted Avg.    0.454    0.454    ?          0.454    ?          ?        0.500     0.307

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3284    0    0    0 |    a = positive
 1293    0    0    0 |    b = negative
 1547    0    0    0 |    c = neutral
 1106    0    0    0 |    d = objective
```

SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3284               45.4219 %
Incorrectly Classified Instances      3946               54.5781 %
Kappa statistic                          0
Mean absolute error                      0.3463
Root mean squared error                  0.4161
Relative absolute error                 99.9949 %
Root relative squared error            100      %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.454      1.000    0.625      ?        0.499     0.454     positive
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.179     negative
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.214     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.153     objective
Weighted Avg.    0.454    0.454    ?          0.454    ?          ?        0.499     0.307

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3284    0    0    0 |    a = positive
 1293    0    0    0 |    b = negative
 1547    0    0    0 |    c = neutral
 1106    0    0    0 |    d = objective
```

- **Second Experiment**

Resource used: Tweepy and TextBlob python package
The API reference can be found at:
http://textblob.readthedocs.io/en/dev/api_reference.html
Python Script: generate.py, sentiment.py
Output folder is ‘E2_API_polarity_result’.

The result from experiment 1 was not acceptable. Though processing time was short, we had a lower accuracy. The script sentiment.py imports two well-encapsulated python packages and it calculates a sentiment polarity score of each message. For this test, we only had the polarity scores as feature.

J48

```
Number of Leaves  :     14

Size of the tree :      27


Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3438               47.5519 %
Incorrectly Classified Instances      3792               52.4481 %
Kappa statistic                          0.0985
Mean absolute error                      0.3265
Root mean squared error                  0.4044
Relative absolute error                 94.2781 %
Root relative squared error             97.1869 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.929    0.821    0.485      0.929    0.637      0.161    0.670     0.589     positive
                 0.283    0.087    0.416      0.283    0.337      0.230    0.658     0.299     negative
                 0.000    0.001    0.000      0.000    0.000      -0.012   0.565     0.247     neutral
                 0.019    0.006    0.368      0.019    0.036      0.053    0.615     0.208     objective
Weighted Avg.    0.476    0.389    0.351      0.476    0.355      0.120    0.637     0.405

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3051  215    3   15 |    a = positive
  919  366    1    7 |    b = negative
 1340  193    0   14 |    c = neutral
  979  106    0   21 |    d = objective
```

SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3284               45.4219 %
Incorrectly Classified Instances      3946               54.5781 %
Kappa statistic                          0
Mean absolute error                      0.3349
Root mean squared error                  0.4268
Relative absolute error                 96.728  %
Root relative squared error            102.5774 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.454      1.000    0.625      ?        0.500     0.454     positive
                 0.000    0.000    ?          0.000    ?          ?        0.539     0.209     negative
                 0.000    0.000    ?          0.000    ?          ?        0.503     0.215     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.153     objective
Weighted Avg.    0.454    0.454    ?          0.454    ?          ?        0.508     0.313

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3284    0    0    0 |    a = positive
 1293    0    0    0 |    b = negative
 1547    0    0    0 |    c = neutral
 1106    0    0    0 |    d = objective
```

NB

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3284               45.4219 %
Incorrectly Classified Instances      3946               54.5781 %
Kappa statistic                          0
Mean absolute error                      0.3463
Root mean squared error                  0.4161
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.454      1.000    0.625      ?        0.499     0.454     positive
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.179     negative
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.214     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.499     0.153     objective
Weighted Avg.    0.454    0.454    ?          0.454    ?          ?        0.499     0.307

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3284    0    0    0 |    a = positive
 1293    0    0    0 |    b = negative
 1547    0    0    0 |    c = neutral
 1106    0    0    0 |    d = objective
```

- **Third Experiment**

On top of the first and second experiment .arff data files, we added and combined the features. New features includes a weight of a message; it was calculated by generate.py. Based on ‘TaboadaGrieve2004-SO.csv’, we accumulated each word’s weight; the more positive a message is the higher weight it gets. Overall, we had positive words count, negative words count, weight, polarity, and if contains question mark as features. We had a higher accuracy of J48.

J48

```
Number of Leaves  :     221

Size of the tree :      441


Time taken to build model: 0.11 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3500               48.4094 %
Incorrectly Classified Instances      3730               51.5906 %
Kappa statistic                          0.1578
Mean absolute error                      0.3112
Root mean squared error                  0.407
Relative absolute error                 89.883  %
Root relative squared error             97.8275 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.854    0.663    0.518      0.854    0.645      0.220    0.701     0.623     positive
                 0.252    0.061    0.475      0.252    0.329      0.250    0.647     0.318     negative
                 0.154    0.085    0.331      0.154    0.211      0.095    0.595     0.268     neutral
                 0.117    0.044    0.322      0.117    0.171      0.114    0.657     0.245     objective
Weighted Avg.    0.484    0.337    0.440      0.484    0.423      0.183    0.662     0.435

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2806  153  209  116 |    a = positive
  769  326  141   57 |    b = negative
 1078  131  239   99 |    c = neutral
  769   76  132  129 |    d = objective
```

SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3284               45.4219 %
Incorrectly Classified Instances      3946               54.5781 %
Kappa statistic                          0
Mean absolute error                      0.333
Root mean squared error                  0.4245
Relative absolute error                 96.1555 %
Root relative squared error            102.0177 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.454      1.000    0.625      ?        0.500     0.454     positive
                 0.000    0.000    ?          0.000    ?          ?        0.607     0.241     negative
                 0.000    0.000    ?          0.000    ?          ?        0.505     0.216     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.520     0.159     objective
Weighted Avg.    0.454    0.454    ?          0.454    ?          ?        0.523     0.320

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3284    0    0    0 |    a = positive
 1293    0    0    0 |    b = negative
 1547    0    0    0 |    c = neutral
 1106    0    0    0 |    d = objective
```

NB

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3233               44.7165 %
Incorrectly Classified Instances      3997               55.2835 %
Kappa statistic                          0.2115
Mean absolute error                      0.3055
Root mean squared error                  0.4084
Relative absolute error                 88.2162 %
Root relative squared error             98.1463 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.640    0.302    0.638      0.640    0.639      0.337    0.715     0.685     positive
                 0.187    0.042    0.490      0.187    0.271      0.220    0.666     0.345     negative
                 0.165    0.099    0.313      0.165    0.217      0.086    0.623     0.288     neutral
                 0.573    0.325    0.242      0.573    0.340      0.186    0.674     0.237     objective
Weighted Avg.    0.447    0.216    0.481      0.447    0.437      0.239    0.680     0.470

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2101  109  254  820 |    a = positive
  398  242  167  486 |    b = negative
  513   93  256  685 |    c = neutral
  282   50  140  634 |    d = objective
```

- **Fourth Experiment**

Since the highest accuracy so far has not reached beyond 50%. We were thinking the reasons that restrains the accuracy. Reasons we concluded:

- The features we were using were not strong enough to classify a message to one of the four categories. Since so far we only have features tell us if a message tends to approach positive or negative. Features, such as positive/ negative word number, weight, and polarity score, only give information of which direction a message might be approaching to. However, we have four directions that a message can approach. Thus, the vector space might be skewed.

This time we tried a new set of features includes the positive sentiment score, neutral score, negative score, and a compound score. We made use of a existing well encapsulated package ‘nltk.sentiment’ package. The classification results of three classifiers has improved and reached beyond 50 percent.

J48

```
Number of Leaves  :     11

Size of the tree :      21


Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3815               52.7663 %
Incorrectly Classified Instances      3415               47.2337 %
Kappa statistic                          0.309
Mean absolute error                      0.295
Root mean squared error                  0.3882
Relative absolute error                 85.1848 %
Root relative squared error             93.2879 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.717    0.255    0.701      0.717    0.709      0.461    0.738     0.654     positive
                 0.542    0.129    0.478      0.542    0.508      0.394    0.735     0.404     negative
                 0.450    0.259    0.321      0.450    0.375      0.171    0.617     0.282     neutral
                 0.058    0.029    0.268      0.058    0.095      0.059    0.696     0.264     objective
Weighted Avg.    0.528    0.199    0.514      0.528    0.508      0.326    0.705     0.470

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2354  333  531   66 |    a = positive
  293  701  273   26 |    b = negative
  463  305  696   83 |    c = neutral
  250  126  666   64 |    d = objective
```

SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3820               52.8354 %
Incorrectly Classified Instances      3410               47.1646 %
Kappa statistic                          0.2915
Mean absolute error                      0.3183
Root mean squared error                  0.4065
Relative absolute error                 91.9149 %
Root relative squared error             97.6952 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.755    0.332    0.655      0.755    0.701      0.422    0.721     0.621     positive
                 0.462    0.093    0.520      0.462    0.489      0.387    0.729     0.381     negative
                 0.480    0.273    0.324      0.480    0.387      0.183    0.609     0.271     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.637     0.207     objective
Weighted Avg.    0.528    0.226    ?          0.528    ?          ?        0.686     0.440

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2481  243  560    0 |    a = positive
  405  597  291    0 |    b = negative
  589  216  742    0 |    c = neutral
  315   93  698    0 |    d = objective
```

NB

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3779               52.2683 %
Incorrectly Classified Instances      3451               47.7317 %
Kappa statistic                          0.3164
Mean absolute error                      0.2602
Root mean squared error                  0.4059
Relative absolute error                 75.1524 %
Root relative squared error             97.5478 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.722    0.266    0.693      0.722    0.708      0.455    0.770     0.737     positive
                 0.470    0.098    0.510      0.470    0.489      0.384    0.769     0.455     negative
                 0.047    0.030    0.303      0.047    0.082      0.040    0.654     0.297     neutral
                 0.656    0.269    0.306      0.656    0.417      0.297    0.735     0.285     objective
Weighted Avg.    0.523    0.186    0.518      0.523    0.490      0.329    0.740     0.523

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2372  269   46  597 |    a = positive
  310  608   92  283 |    b = negative
  480  224   73  770 |    c = neutral
  259   91   30  726 |    d = objective
```

- **Fifth Experiment (Best accuracy achieved)**

Resource used: NLTK python package
The API reference can be found at:
http://www.nltk.org/api/nltk.sentiment.html#module-nltk.sentiment.sentiment_analyzer
Python Script: generate.py, NLTKAPI.py
Output folder is ‘E5_polarity_sentimnet_result’.

There is no new features included. We used polarity score and compound score calculated by nltk.sentiment package as features. We got the highest accuracy (53.029%).

J48

```
Number of Leaves  :     43

Size of the tree :      85


Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3834               53.029  %
Incorrectly Classified Instances      3396               46.971  %
Kappa statistic                          0.3092
Mean absolute error                      0.294
Root mean squared error                  0.3876
Relative absolute error                 84.9048 %
Root relative squared error             93.1495 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.760    0.322    0.662      0.760    0.708      0.436    0.742     0.658     positive
                 0.513    0.122    0.478      0.513    0.495      0.380    0.737     0.418     negative
                 0.131    0.072    0.332      0.131    0.188      0.087    0.624     0.298     neutral
                 0.427    0.162    0.323      0.427    0.367      0.237    0.695     0.273     objective
Weighted Avg.    0.530    0.208    0.507      0.530    0.506      0.321    0.709     0.479

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2496  306  150  332 |    a = positive
  367  663   96  167 |    b = negative
  565  287  203  492 |    c = neutral
  340  131  163  472 |    d = objective
```

SMO

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3652               50.5118 %
Incorrectly Classified Instances      3578               49.4882 %
Kappa statistic                          0.1638
Mean absolute error                      0.3283
Root mean squared error                  0.4187
Relative absolute error                 94.7942 %
Root relative squared error            100.6198 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.928    0.753    0.506      0.928    0.655      0.233    0.591     0.506     positive
                 0.464    0.099    0.506      0.464    0.484      0.378    0.703     0.356     negative
                 0.004    0.004    0.214      0.004    0.008      0.000    0.496     0.213     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.372     0.133     objective
Weighted Avg.    0.505    0.360    ?          0.505    ?          ?        0.557     0.359

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 3046  227   11    0 |    a = positive
  686  600    7    0 |    b = negative
 1289  252    6    0 |    c = neutral
  996  106    4    0 |    d = objective
```

NB

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3783               52.3237 %
Incorrectly Classified Instances      3447               47.6763 %
Kappa statistic                          0.2924
Mean absolute error                      0.2826
Root mean squared error                  0.3856
Relative absolute error                 81.6015 %
Root relative squared error             92.6711 %
Total Number of Instances             7230

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.770    0.339    0.654      0.770    0.707      0.430    0.779     0.746     positive
                 0.440    0.099    0.491      0.440    0.464      0.356    0.755     0.462     negative
                 0.191    0.115    0.312      0.191    0.237      0.093    0.652     0.294     neutral
                 0.353    0.142    0.310      0.353    0.330      0.200    0.717     0.275     objective
Weighted Avg.    0.523    0.218    0.499      0.523    0.505      0.310    0.738     0.526

=== Confusion Matrix ===

    a    b    c    d   <-- classified as
 2528  229  235  292 |    a = positive
  366  569  184  174 |    b = negative
  608  241  296  402 |    c = neutral
  364  119  233  390 |    d = objective
```

- **Experiment out of Curiosity**

The highest accuracy so far is about 53%. According to statistics probability, the probability that we correctly classify a message to one of the four class labels is 1/4. Therefore, our accuracy is acceptable. We were curious why it could not reach a higher result. First guess would be our calculated sentiment score did not specially consider the 'objective' category. To prove our assumption, we did a sixth experiment.

In this test, when parsing all the messages, we changed 'objective' label to 'neural'. We got a much higher accuracy — 63.444%.

J48 pruned tree
---

compund_score <= 0.3695
| neu_score <= 0.844
| | compund_score <= 0.1551: negative (1158.0/564.0)
| | compund_score > 0.1551: positive (258.0/119.0)
| neu_score > 0.844: neutral (2759.0/1072.0)
compund_score > 0.3695: positive (3055.0/839.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**4587**	**63.444 %**
Incorrectly Classified Instances	**2643**	**36.556 %**
Kappa statistic	0.4149	
Mean absolute error	0.253	
Root mean squared error	0.3571	
Relative absolute error	80.6825 %	
Root relative squared error	90.1975 %	
Total Number of Instances	7230	

SMO
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**4559**	**63.0567 %**
Incorrectly Classified Instances	**2671**	**36.9433 %**
Kappa statistic	0.4025	
Mean absolute error	0.2909	
Root mean squared error	0.3716	
Relative absolute error	92.7736 %	
Root relative squared error	93.8499 %	
Total Number of Instances	7230	

The last experiment was more biased, since we simply changed all 'neutral' label to 'positive'.

J48

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**6005**	**83.0567 %**
Incorrectly Classified Instances	**1225**	**16.9433 %**
Kappa statistic	0.3186	
Mean absolute error	0.1231	
Root mean squared error	0.2485	
Relative absolute error	83.718 %	
Root relative squared error	91.7041 %	
Total Number of Instances	7230	