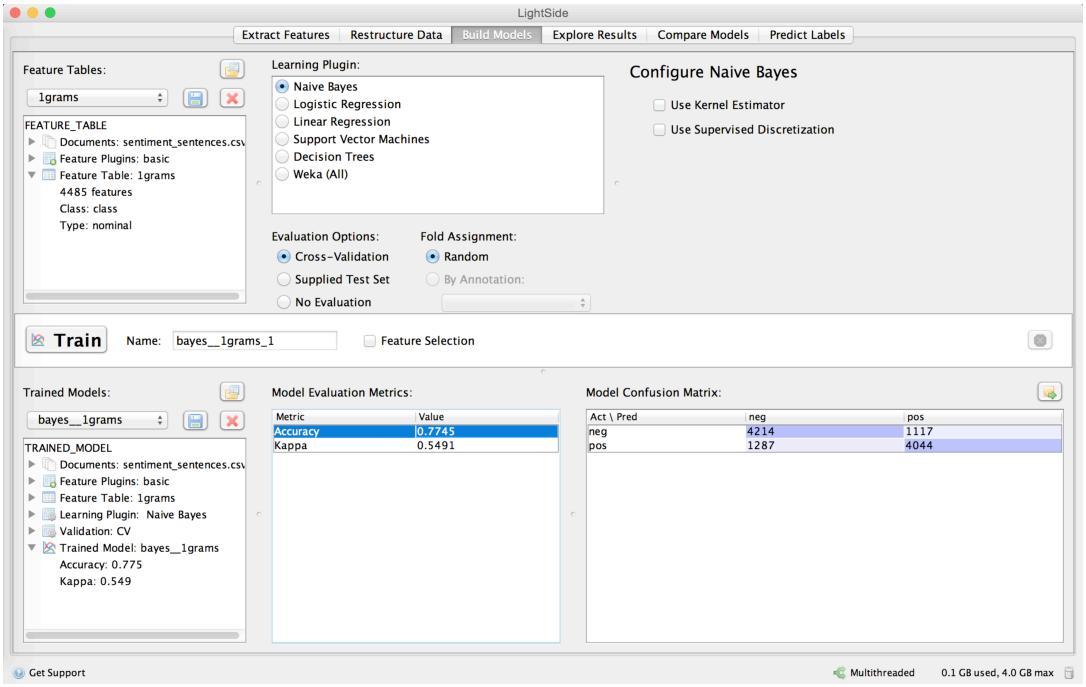
Assignment 5

Jingyuan Liu

Deliverables:

1. List your performance from 5. Turn in your .arff file.
   The result of LightSIDE running on the datasets with unigram as feature space is as follows:



   We can see from the figure that the prediction accuracy is 0.7745, and the Kappa of the model is 0.5491.

2. List your 5 rules here
   a. Rule 1:
      chemistry

   b. Rule 2:
      (romance OR romantic) AND love

   c. Rule 3:
      unnatural OR immaturity

   d. Rule 4:
      shallow OR annoying

   e. Rule 5:
      (NOT intended) OR (unintentional)

3. Fill in the following table:

	Value of Rule 1	Value of Rule 2	Value of Rule 3	Value of Rule 4	Value of Rule 5
Positive Review 1	1	1	0	0	0
Positive Review 2	1	1	0	0	0
Positive Review 3	0	1	0	0	0
Positive Review 4	1	1	0	0	0
Positive Review 5	1	1	0	0	0
Negative Review 1	0	0	1	0	1
Negative Review 2	0	1	0	1	0
Negative Review 3	0	0	1	0	1
Negative Review 4	0	0	0	1	0
Negative Review 5	0	0	0	0	1

4. Indicate which feature you think worked best and what you think made that feature work better than the others.

I think the best features are the RULE 1, RULE 2, and RULE 3. They are better than the other two features.

For the RULE 1, I used chemistry, which always means that actors worked well with each other. In negative reviews, it is not likely to appear this word.

For the RULE 2, all the positive reviewers express their love for the movie. However, the negative also appeared this word in the phrase “True Love”. Therefore, combining a “AND” with the “romantic OR romance”, which share the meaning but in different forms, would improve the classification performance of the feature.

For the RULE 3, not intended is the same meaning of unintentional. Negative reviewers would use this word to criticize the actors.