Performance difference table:

Naive Bayes:

Attributes	precision	recall	accuracy	F-measure
a1	-0.014	-0.02099999999999	-2.12%	-0.026
a2	0.02	-0.01299999999999	-1.34%	-0.014
a3	0.032	0.024	2.42%	0.031
a4	0.294	0.222	22.22%	0.242
a5	0.046	0.025	2.49999999999999	0.028
a6	0.013	-0.00900000000000	-0.90000000000000	-0.00600000000000
a7	0.001	-0.02099999999999	-2.08%	-0.024
baseline (actual value, not difference)	0.637	0.553	55.28%	0.506

Decision Tree:

Attributes	precision	recall	accuracy	F-measure
a1	0	0	0%	0
a2	0	0	0%	0
a3	0	0	0%	0
a4	0.564	0.439	43.92%	0.47
a5	0	0	0%	0
a6	0	0	0%	0
a7	0.001	0	0.039999999999995	0
baseline (actual value, not difference)	0.863	0.711	71.1%	0.710

Question 1:

Which is the best machine learning model (classifier) for this task? You need to discuss this per metric used to compute the performance.

Based on the baseline system in both performance difference tables, I believe the decision tree model has the best performance. Because its result has highest accuracy, highest precision and highest recall.

Question 2:

> Which features contributed the most to the performance (precision, recall, accuracy, and F-measure)? Which contributed the least? (you need to do this for each machine learning model considered here)

Based on both performance difference tables, we can see that a4 contributes the most to the performance because it creates most difference on all evaluation parameters in both tables when I remove it by "leave one out" approach.

In the scenario of the Naive Bayes classifier, I think a6 contributes the least. Even though it doesn't make least difference on precision, it makes the least contribution on recall, accuracy and F-measure on Naive Bayes model.

In the scenario of the decision tree model, all attributes except a4 make zero contribution. So I believe that these attributes except a4 all contribute least. Another thing worth mentioning is that there might be a rounding or calculating error on a7, which leads to slight change on precision and accuracy. But I think it won't influence my conclusion on human judgement.

Question 3:

> How good is this feature set (i.e., the seven attributes) for this task? (based on the answer at Question 2)

Based on the results on question 2, I believe that a4 is the most informative because it is the target word for part-of-speech tagging, which quite makes sense.

According to the result from the Naive Bayes Classifier, I believe that there is a trend that the further distance between words and the target word, the less relevant they are.

According to the result from the Decision Tree Classifier, I think the only worthwhile attribute is a4, and others make no contribution on.

Question 4:

> If you had to solve the problem in a different (and probably more efficient way), which attributes would you choose?

In order to solve the problem efficiently, I believe we need to shrink the effective size of our dataset and then get more precise result by eliminating the intervention of unrelated words.

Based on my answer about the Naive Bayes model on question 3, I will try to remove the a1 and a7, because both are too far from the target word a4 and not relevant very much.

By running the WEKA system, we can see that the system performance improves by leaving a1 and a7 out.

The performance table for the Naive Bayes model: (actual value, not difference)

Attributes	precision	recall	accuracy	F-measure
leave a1 and a7	0.668	0.599	59.88%	0.562
Baseline	0.637	0.553	55.28%	0.506

Extra Credit:

See README.md file.