Dataset:

A: Student Answer

R\_1: Reference answer 1

R\_2: Reference answer 2

R\_3: Refrence answer 3

L: Correct

A: Student Answer

R\_1: Reference answer 1

R\_2: Reference answer 2

R\_3: Refrence answer 3

L: Correct-but-incomplete

Approach

1. Convert A, R\_1, R\_2, R\_3 into vector with doc2vec

- Say D\_A, D\_R\_1, D\_R\_2, D\_R\_3 are obtained doc2vec

2. Calculate similarity score between (D\_A, D\_R\_1), (D\_A,D\_R\_2), (D\_A,D\_R\_3)

- I am not sure what similarity measure to use?

- Cosine Similarity score?

3. Save the highest similarity score

4. Repeat the process for all A and R

5. Use similarity sore as future to Multinomial Logistic Regression (in Weka) to predict the

correctness label