

Gold label

non-derog.

0.84

0.16

derogatory

0.045

0.96

non-derog.

derogatory

Annotator label

