1. Download Ubuntu corpus dataset from "*https://s3.amazonaws.com/ngv-public/data.zip*"

2. Install python 3.6 and install requirements from requirements.txt

* *pip install -r requirements.txt*
* *source activate ml*

3. Update the path variables with links to the data and where you want to save model output in main.py

* *parser.add\_argument('--root\_dir', default='')*
* *parser.add\_argument('--dataset\_train\_path', default='')*
* *parser.add\_argument('--dataset\_test\_path', default='')*
* *parser.add\_argument('--dataset\_val\_path', default='')*
* *parser.add\_argument('--vocab\_path', default='')*
* *parser.add\_argument('--model\_save\_dir', default='')*
* *parser.add\_argument('--test\_tube\_dir', default='')*

4. Run the code for training and testing

* *python main.py*