In response to latest feedback:

3 – “You indicated you are building an extractive summarizer. How do you test its performance? If it is difficult to do, what can you use as a starting point? Each project must have a way to test the underlying hypothesis of the project.”

This is an **unsupervised** project. There are no p-values or confidence intervals to report. The only metric I could find is ROGUE – could not get it implemented because so many dependencies changed. I know nothing about PERL.

Regardless, we – humans – are the gold standard. One of the difficulties and expenses in real projects of this nature is getting data that has been hand evaluated be it for summarization, sentiment analysis, etc.

So, all you can do in a project at **this level** is read the summaries and pass your own judgement. My judgement is that BERT was the best. In my presentation I explained why BERT works so well – it is context aware. It knows (not the same as people ‘know’) that the word LEFT used in LEFT BANK and LEFT HAND are not the same and vectorizes appropriately.

4 – “feedback given my previous grader still holds. Provide a clearly labeled conclusion section at the end of the python notebook.”

My conclusion is basically the one shown in the Google slides:

**Based on this limited example, BERT is by far the best summarizer.**

**But -- BERT is slow unless you have a lot of TPUs (Tensor Processing Units)**

**This is for a Textrank based extractive summarizer.**

**Implementing this as a production ready abstractive summarizer has not been done.**

**Once one is built, the true power of BERT will be unleashed.**

**Until then, human rating is still the gold standard.**

I started out planning to build a BERT based abstractive summarizer – then I discovered that no one had done it. It is an appropriate project for someone like our specialization’s instructor, Dennis Tran. **Or maybe one of the graders**. Or Google Brain.

Again, when I stated BERT is by far the best summarizer, it is based on my reading the summaries. The Gold Standard.

2- “Shows count vectorizer frequency table, but it is not readable. One suggestion is choosing a top K to show.”

**That is exactly what a TEXTRANK based summary shows!**