*Notes.*

**Models:**

1. pretraining BERT from huggingface (bert\_test.ipynb and bert\_logs folder)

2) Simple LSTM (lstm\_test.ipynb and lstm\_logs folder)

**Exp types:**

I used 3 experiment types.   
“initial” - means what model trained without filter “ne ok” calmdown ex.

“withoutneok” - model trained on filtered “ne ok” calmdown ex.

“random binary” - trained on random labels. (just for fun)

**Counts:**

Total number of items 655

Number of categories

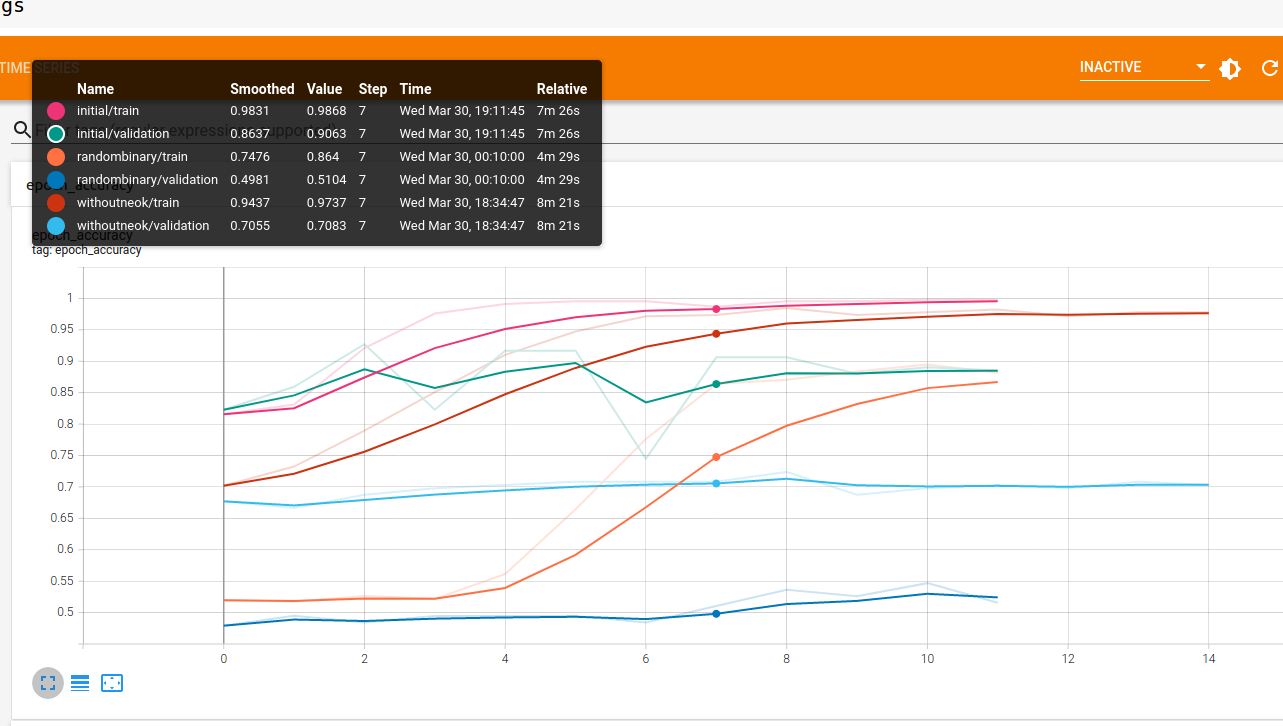
calmdown ex 537 (other exercies 79, and 39)

after filtering ne ok calmdown exerices

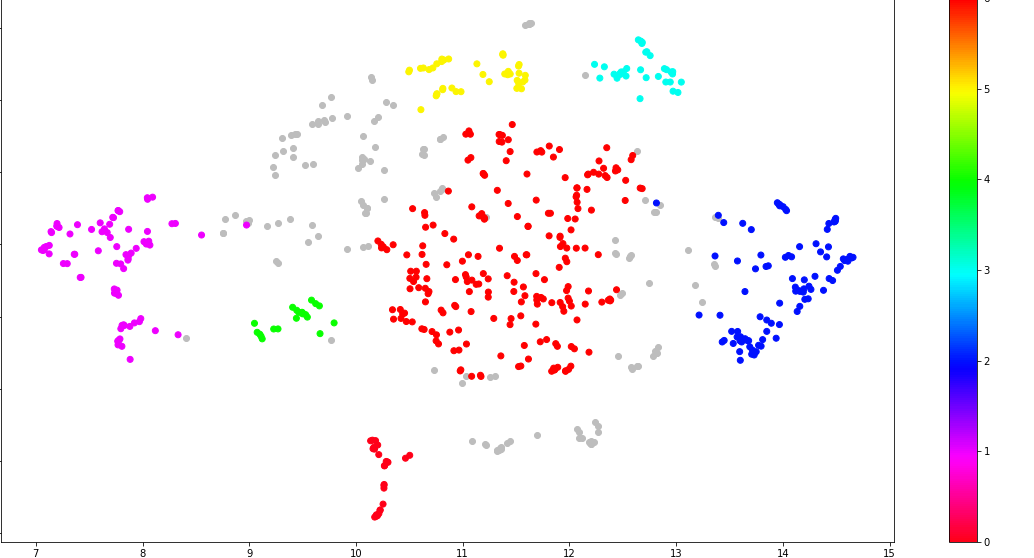
Num of calmdown ex 186

Num of other 469

some BERT accuracy training results … ( I can notice what these results are comparable with LSTM’s)



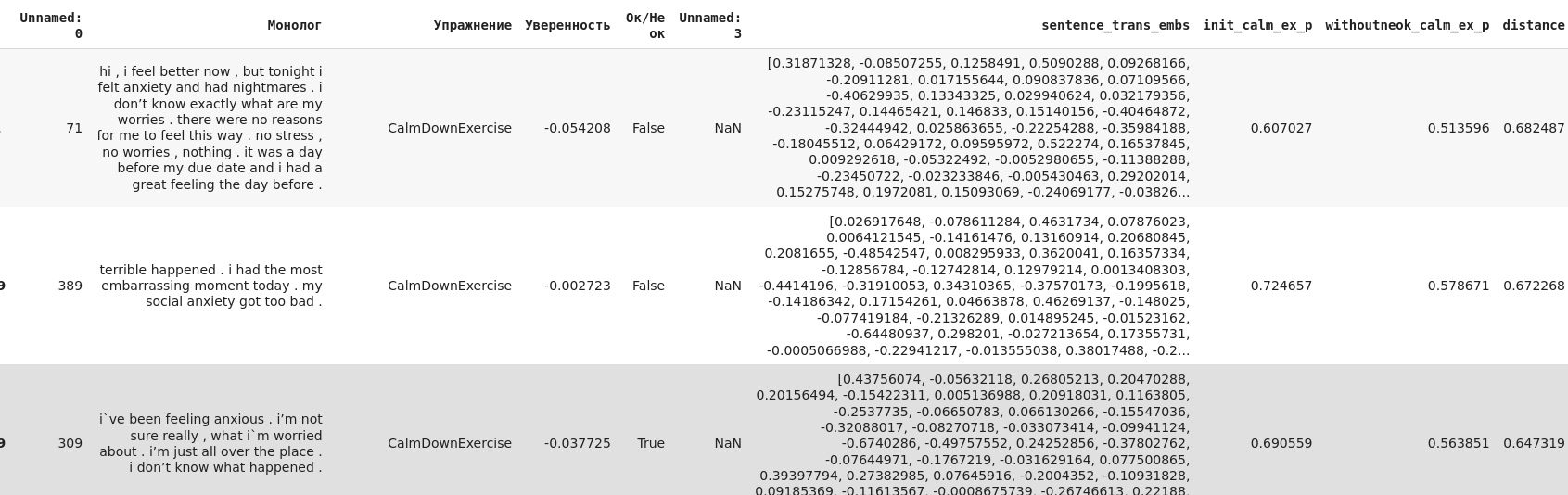
**Analysis.ipynb**



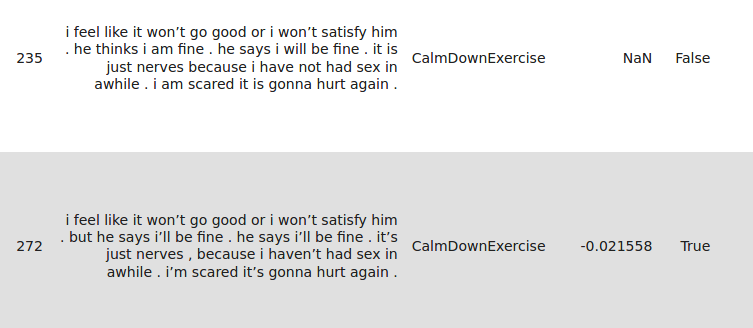
Example of show nearest and result.csv (which in this folder)  
init\_calm\_ex\_p - prediciton from BERT trained on unfiltered ok/neok data

without\_neok\_calm\_ex\_p - prediction from BERT trainined on without neok data..

distance - distance to selected item…



FYI, I understand correctly some examples can be repeatable (and also can have different ok/neok)



**Visualization via tensorboard projector.**

use embeddings.tsv and metadata.tsv or metadata2.tsv (with labels) to upload for projector

<https://projector.tensorflow.org/>

https://amitness.com/interactive-sentence-embeddings/

