Datasets need to be uploaded/called for the .ipynb files to run.

* **Datasets:**

1. **Original Dataset with no emojis**
   1. emotions1 - emotions1.csv - For Baselines, CNN and LSTM
   2. emotions1 - emotions1.xlsx - For BERT and XLNET
2. **Dataset with Emojis as is**
   1. Final\_Emojis - emotions1.csv - For Baselines, CNN and LSTM
   2. Final\_Emojis - emotions1.xlsx - For BERT and XLNET
3. **Dataset with Emojis converted to words**
   1. Dataset\_With\_Emoji.csv - For Baselines, CNN and LSTM

* **Models:**

1. **Baseline Models**

- Both1a and 3a should be present locally to get the output of the Baseline models with and without emojis.

* 1. **Baselines** - Baseline.ipynb

1. **CNN**

- Both 1a and 3a should be present locally to get the output of CNN with and without emojis.

* 1. **CNN without using word2vec -** CNN with word2vec.ipynb
  2. **CNN using word2vec -** CNN with word2vec.ipynb

1. **LSTM**

- The ipynb file is called LSTM.ipynb (this includes all combinations of LSTM embedding layers and is also the file that contains Word2vec+Emoji2vec on CNN) ; Files to be present locally are

- Emoji2vec.bin

- Phrase2vec.py

- The 3 datasets mentioned above

1. **BERT**

* By default emoticons to words mapping is enabled. To disable it, comment cell 11 (apply(token\_emoji\_conversion) method).
* Mount the drive (helpful in saving checkpoints too) and make sure that dataset is at “/content/gdrive/My Drive/Final\_Emojis - emotions1.csv” location.
* Use GPU for code execution
* Specify checkpoint directory under - OUTPUT\_DIR in code
  1. **BERT for cased -** BERT\_converting emoji to word\_cased.ipynb; the generated probabilities are saved on the drive in a .csv file named prob\_uncased.csv
  2. **BERT for uncased -** BERT\_converting emoji to word\_uncased.ipynb; the generated probabilities are saved on the drive in a .csv file named prob\_cased.csv
  3. **BERT(cased+uncased) -** read cased and uncased csv files in variables cased\_prob and uncased\_prob in the code.

1. **XLNET**

* Mount the drive and make sure that dataset is at the “/content/gdrive/My Drive/Final\_Emojis - emotions1.csv” location.
* Specify path from which checkpoint is to be loaded in torch.load
* Use GPU for code execution
* Follow instructions specified in XLNet.pynb notebook.
  1. **For XLNet** - XLNet.ipynb