(1) Self-supervised Learning Step

Model: BERT

Dataset:

BooksCorpus

+

Wikipedia

Objectives: Predict - the masked words (MLM)
- the next sentences (NSP)

(2) Supervised Learning Step

Model:
(pre-trained from step #1)

Dataset:

Email messages	Class
Buy these pills…	Spam
Win cash prizes…	Spam
Dear Mr. Dupont, please find…	Not Spam