Aukeratu lehenik BERT eta finetune

Gero aukeratu BIONER edo BIOPUBNER edo holakoaketa probatu eta finetune ikusteko nola hobetzen duten

Seguruaski behar izango dugula BIO etiketatzea

Hasieran etiketatu maizen agertzen direnak

Gero probatu nire interesatzen zaizkidanak

umls bert

<https://huggingface.co/microsoft/BiomedNLP-PubMedBERT-base-uncased-abstract-fulltext>

<https://arxiv.org/pdf/2007.15779.pdf>

<https://towardsdatascience.com/whos-who-and-what-s-what-advances-in-biomedical-named-entity-recognition-bioner-c42a3f63334c>

<https://arxiv.org/pdf/2001.08904.pdf>

ZALANTZAK:

* ondo evaluate jartzea evaluateF1 eta accuracy izatea F1?
* zer egiten dut ez baldin badut etiketa hori nire etiketetan? Jartzen diot O? edo jartzen diot B-MED? O dauka zentzu gehio
* zer geratzen de dev eta test-en etiketa berri bat badago trainen ez dagoena? errorea ematen dit, zer egiten dut? etiketa berriak sortu? edo..?
* 2 etiekta dituenean igual da hobe banatuta egitea, horrela igual bat asmatzen du bat ere ez asmatu beharrean
* txarrago egiteak zentsua dauka ezta?
* zeba ematen dit loss altua baina emaitz onak?