

BERT-NER WorkFlow

BERT TRAINING---

Have proper Manually Tagged Data (List of strings)

1. **CLASS_BIO_Format_Tagging** --> This prepares BERT Training Format Data (BIO)
2. **CLASS_BERT_TRAINING** --> Used Data received from 1 in this (Model saving inbuilt)
Model saving done here itself

New Data Prediction & Accuracy Metrics SHEET Generations---

1. Load the saved model from point 2, and run **CLASS_BERT_PREDS_DATAFRAME** file for generating BERT Tags on the required Text. [Also have the functionality to keep english predictions of BERT only (if needed in future)]
2. **CLASS_All_Accuracy_sheet_GENERATOR**(Both notebook and Abstraction file available) --> Used for Accuracy Sheet generation(keep Technology words Dictionary and English words Dictionary ready) for three criterias, keeping Tags from English Dictionary only(got from BERT), keeping Tag from Technology Dictionary(got using String Match) & Merging these two as final sheet which gives combined Accuracy stats on English and Technology Dictionary.

[*All code blocks have been verified by running them on our data (as on 24/07/21)]