**Data Augmentation**

**Data Augmentation Technique** is a process that enables us to artificially increase training data size by generating different versions of real datasets. is It used in Natural Language Processing (NLP) to deal with data scarcity and insufficient data diversity. Data augmentation increases the training data size, which improves the model’s performance. The distribution of augmented data generated should neither be too similar nor too different from the original. This may lead to overfitting or poor performance through Effective DA approaches should aim for a balance.

Hence, we have used **generative adversarial network (GAN)** which is trained to generate text with a few words and generative language models like **BERT**, RoBERTa, BART and T5 model can be used to generate the text in a more class category preserving manner. The generative model encodes the class category along with its associated text sequences to generate newer samples with some modifications. This approach is usually more reliable and the sample generated is more representative of the associated class category.

Library used : ‘Textattack’ → WordNetAugmenter



