

Identification of structures in Disclosure text using BERT model: A Transformer Based Approach

Objective:

- To Understand Preprocessing for NLP downstream tasks
- To Understand Use of pre-trained models and embedding for NLP downstream tasks
- To Understand BERT architecture and Use
- To Understand essence of *Transfer Learning*

Dataset:

Download the Students' essays dataset from the given Kaggle [link](#). The dataset is publicly available and is part of Kaggle Competition.

Assignment 1. Data Preprocessing

- Handle typo errors
- Handle Null values and duplicates
- Do feature selection
- Do Label Encoding
- Split the data into train test and validation using *train test split* from *Sk-learn library*.

Assignment 2. Training BERT Model and Transfer learning

- Load a pre-trained BERT model (Preprocessing and embeddings) from *Keras Hub*.
- Design model according to your requirements (Dropout layer, number of neurons in last layer, Activation function).
- Compile the model (Select appropriate optimizer, loss functions and evaluation metrics)
- Transfer learning: train the model on your dataset.

Assignment 3. Results

- Plot different graphs for specified epochs like training accuracy vs training loss, validation accuracy vs validation loss, training accuracy vs validation accuracy.
- Plot confusion matrix to understand the results better.
- Print classification report specifying precision and recall.