

CSE440: Natural Language Processing II

Lab Assignment 4

Task: AG News Classification Dataset

- Download and Load [AG News Classification Dataset](#) from kaggle.
- Perform all necessary Exploratory Data Analysis (EDA):
 - Check dataset size, number of classes, and presence of missing values
 - Display class distribution using a count plot or bar chart
 - Analyze text statistics such as average sentence length, word frequency, etc
- Carry out all required Data Preprocessing:
 - Tokenize the text and pad sequences
 - Split the dataset into training and test sets
 - If the dataset is imbalanced, apply balancing techniques such as Downsampling (reducing majority classes) or Upsampling / Data augmentation (increasing minority classes)
- Train two separate models LSTM (BiLSTM) and GRU (BiGRU) on the training set. You may choose hyperparameters based on your preference but have to justify your choices.
- Evaluate the models using Classification Report and Confusion Matrix. Compare the results and briefly discuss which model performed better and why might that be the case?