Assignment 6

Problem Statement

Sentence Prediction using Phonemes

Data:

https://www.kaggle.com/datasets/mfekadu/darpa-timit-acousticphonetic-continuous-speech

Task:

1. Use DR1 and DR2 dialect for training the model and DR3 dialect for evaluating the model.
2. Pick only .PHN, .WRD and .TXT files for each specified dialect distribution.
3. Create a vocabulary for each word with its associated phonemes.
4. Train the model with available vocabulary for phoneme to word prediction. (Use sequence modeling).
5. Create a pandas dataframe containing the sentence phonemes for each file.
6. Predict the words of all phonemes present in the sentence using the trained model.
7. Calculate the distance between predicted and actual words for each sentence.
8. Apply the same process in the testing set, provide your inferences.
9. Publish the solution notebook in that Kaggle dataset.