Venkata Ashita Medam
SE21UCSE243

NLP Assignment - 3

This repository contains code for sentiment analysis using Word2Vec and FastText word embeddings trained with an RNN model. The models are trained on the IMDB movie review dataset.

Directory Structure :

```
├─── data
|    ├─── train_data.csv
|    └─── test_data.csv
|    └─── train (folder)
|    └─── test (folder)
├─── models
|    ├─── word2vec_rnn_model.pth
|    └─── fasttext_rnn_model.pth
|    ├─── word2vec model
|    └─── fasttext model
```

GitHub Public Access link : https://github.com/ashitamv/Training-a-RNN-model-on-the-given-dataset/tree/main

Execution Instructions :

- Open the ‘se21ucse243.ipynb’ notebook in Google Colab.
- Mount your Google Drive to access the train and test data folders.
- Execute the cells in the notebook sequentially to load the data, preprocess it, train Word2Vec and FastText models, train the RNN model, and evaluate the models.

Implementation Details :

- Language: Python
- Framework: PyTorch
- Word Embeddings: Trained using Gensim's Word2Vec and FastText models.
- RNN Model: Consists of an embedding layer, followed by a GRU layer, and a fully connected layer.
- Training: Cross-entropy loss and the Adam optimizer are used for training.

Analysis and Observations :

- Performance of the Word2Vec and FastText models is evaluated using precision, recall, F1-score, and accuracy metrics.
- Confusion matrices are plotted to visualize the performance of the models on different sentiment classes.
- Both models achieve high accuracy on the sentiment classification task.