Sentiment Analysis - Machine Learning and Deep Learning Models

Section 1: Machine Learning Models
This section contains machine learning models for sentiment analysis, including logistic regression and SVM. The models are trained and evaluated using cross-validation techniques, with performance metrics such as accuracy, precision, recall, and F1-score.

Section 2: Deep Learning Models
Imports and Dataset Preparation

- **Libraries**: Uses PyTorch for deep learning, NumPy for numerical operations, and scikit-learn for evaluation.
- **Custom Dataset**: SentimentDataset converts input data into PyTorch tensors and handles sparse matrix transformations.
- **DataLoader Function**: create_data_loaders creates training and test data loaders with batch processing.

Convolutional Neural Network (CNN) Model

- **Architecture**:
  - 1D convolutional layer with 128 filters and kernel size of 5.
  - Max-pooling layer to reduce dimensions.
  - Fully connected layer to map extracted features to output classes.
- **Training**:
  - Uses cross-entropy loss and Adam optimizer.
  - Runs for 5 epochs per cross-validation fold.
  - Prints loss per epoch.
- **Evaluation**:
  - Computes accuracy, precision, recall, F1-score, and AUC.
  - ROC curve is plotted using false positive and true positive rates.

Multi-Layer Perceptron (MLP) Model

- **Architecture**:
  - Three fully connected layers (128, 64, output size).
  - ReLU activation for hidden layers.
  - Softmax activation for multi-class classification.
- **Training and Evaluation**:
  - Uses the same training and evaluation strategy as CNN.
  - Cross-validation performed over 5 folds.

Model Comparison

- **Metrics Comparison**:
  - Accuracy, precision, recall, F1-score, and AUC stored and averaged across folds.
  - Results displayed in a Pandas DataFrame.
- **Visualization**:
  - Bar chart comparing CNN and MLP model performance.
  - ROC curve comparison for both models.

Final Notes
This script implements both machine learning and deep learning approaches for sentiment analysis, comparing their performance using various metrics and visualizations.