Line(s) Explanation

1   Install necessary libraries: This installs the required libraries for the code, including transformers, pandas, scikit-learn, and torch.

4-15   Import libraries: This imports the necessary Python libraries for data manipulation, machine learning, deep learning, and visualization.

18-19   Mount Google Drive: This mounts Google Drive to the Colab environment to access files stored on Google Drive.

22-28   Generate embeddings with RoBERTa: Defines a function to generate embeddings from text using the RoBERTa model. It tokenizes the text, runs it through the model, and computes the mean of the output states to get the embedding.

31-38   Define a Dataset class for PyTorch: Creates a custom Dataset class to handle the data in a format suitable for PyTorch, enabling easy data loading during training and validation.

41-52   Define a simple neural network: Defines a simple feedforward neural network with two hidden layers and dropout for regularization, used for classifying the text embeddings into real or fake news.

55-83   Train the model: Defines a function to train the neural network. It initializes the data loaders, model, loss function, and optimizer, and trains the model while monitoring validation loss for early stopping.

86-89   Load data from Google Drive: Reads the dataset from a CSV file stored on Google Drive.

92-93   Initialize RoBERTa for embeddings: Loads the tokenizer and model for RoBERTa from the transformers library.

96-98   Prepare data: Converts the text data to a numpy array and the labels to binary format (0 for real, 1 for fake).

101-102   Split the data: Splits the dataset into training, validation, and test sets using train_test_split from scikit-learn.

105-107   Generate embeddings: Uses the generate_embeddings function to create embeddings for the training, validation, and test sets.

110   Define the embedding size: Specifies the size of the embeddings generated by RoBERTa, which is 768.

113-115   Train the model: Calls the train_model function to train the neural network on the training and validation embeddings.

118-125   Evaluate the model: Evaluates the trained model on the test set by predicting labels and calculating accuracy, classification report, and confusion matrix.

128-137   Visualize the confusion matrix: Plots the confusion matrix using Seaborn to visually inspect the model's performance.

140-142   Save the trained model: Saves the trained model's state dictionary to Google Drive for future use.