Fraud Email Detection using BERT - Module Explanation (Text-Only)

Project Objective
The goal of this project is to build a machine learning model that can acc

Dataset & Inputs
The dataset used in this project consists of emails labeled as either frau

Data Preprocessing
Before feeding the data into the BERT model, a series of preprocessing ste
1. Handling Missing Values
2. Label Normalization
3. Data Splitting
4. Conversion to Strings

These steps ensure the data is clean and ready for tokenization and model

Tokenization & BERT Embedding
1. Tokenizer Loading - Load the tokenizer for `bert-base-uncased`
2. Encoding - Convert text into BERT format (`input_ids`, `attention_masks
3. Padding and Truncation - Ensure uniform sequence length
4. Tensor Conversion - Transform data into tensors for PyTorch

Model Design & Training
1. Base Model - Use `bert-base-uncased`
2. Classification Head - A neural network on top of BERT for binary classi
3. Loss Function - Binary Cross-Entropy
4. Optimizer - AdamW with a learning rate scheduler

The model is trained on the training dataset to learn patterns of fraudule

Evaluation & Results
Evaluation is done using:
- Accuracy
- Confusion Matrix
- Precision & Recall
- F1 Score

These metrics help understand how well the model generalizes to unseen ema

Conclusion
This project successfully applies BERT for email fraud detection. BERT's c