**requirements.txt**

pandas

scikit-learn

nltk

**How It Works**

1. **Dataset Loading**
   * The system uses a spam dataset (e.g., CSV file) containing labeled emails (spam or ham).
   * Each row contains the email text and its category.
2. **Text Preprocessing**
   * Converts text to lowercase.
   * Removes punctuation and special characters.
   * Tokenizes into words.
   * Removes common stopwords (e.g., *the, is, and*).
3. **Feature Extraction**
   * Uses **Bag of Words** or **TF-IDF Vectorization** to convert text into numerical vectors.
4. **Model Training**
   * Applies the **Naive Bayes Classifier** (MultinomialNB from scikit-learn) to learn patterns from training data.
5. **Prediction**
   * When a new email is entered, it is preprocessed, vectorized, and passed to the trained model.
   * The model predicts **Spam** or **Not Spam** instantly.
6. **Offline Operation**
   * Once trained, the model runs locally and can classify emails without internet.

✅ **Example Flow**

Email: "Win a brand new iPhone! Click here to claim"

↓

Lowercasing, tokenization, stopword removal

↓

Vectorization into numerical form

↓

Prediction → "Spam"