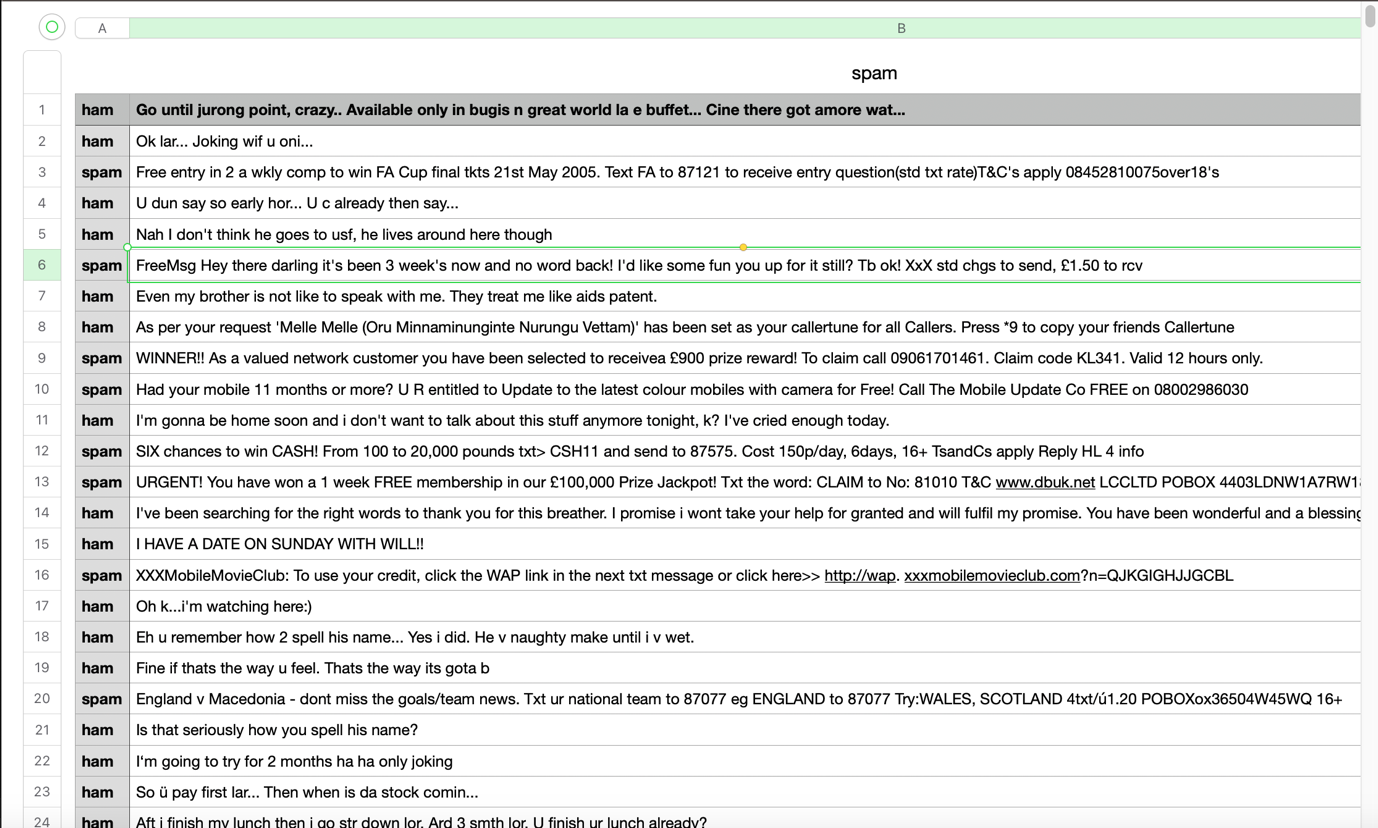
**Building a Smarter AI-Powered Spam Classifier: Designing a Web Application to Classify Spam Messages Using TF-IDF, Multinomial Naive Bayes, and Other NLTK Libraries with Iterative Improvement to Enhance Accuracy, Precision, Recall, and F1-score.**

1. **Data Collection:**

-Download the "spam.csv" dataset from Kaggle's SMS Spam Collection Dataset (https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset).



**2.Data Preprocessing:**

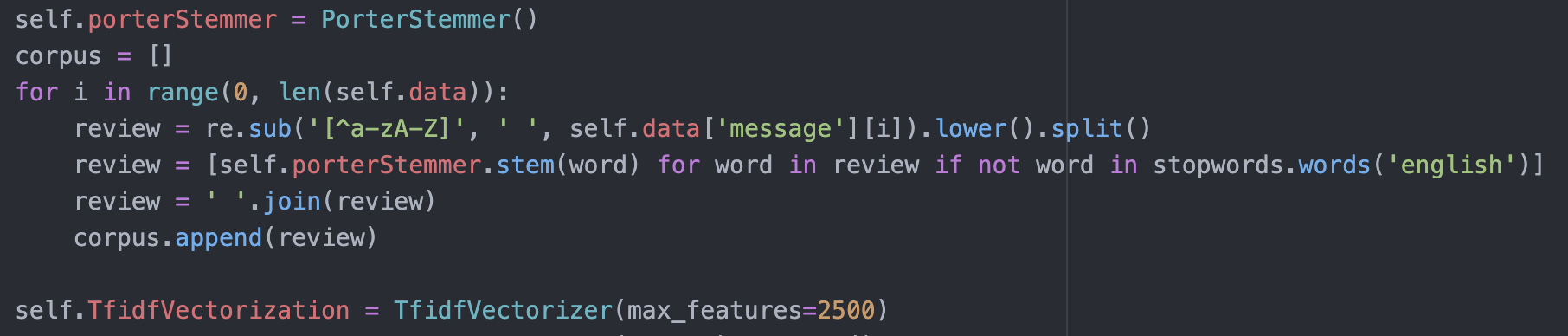
-Load the dataset from "./src/data/spam.csv."

data = pd.read\_csv(r"./src/data/spam.csv",sep="\t",names=["label", "message"])

**3. Feature Extraction:**

- Apply TF-IDF (Term Frequency-Inverse Document Frequency) to convert the tokenized words into numerical features.

- Set the maximum number of features to 2500, meaning you'll create a TF-IDF matrix with 2500 columns representing the most important words in your dataset.

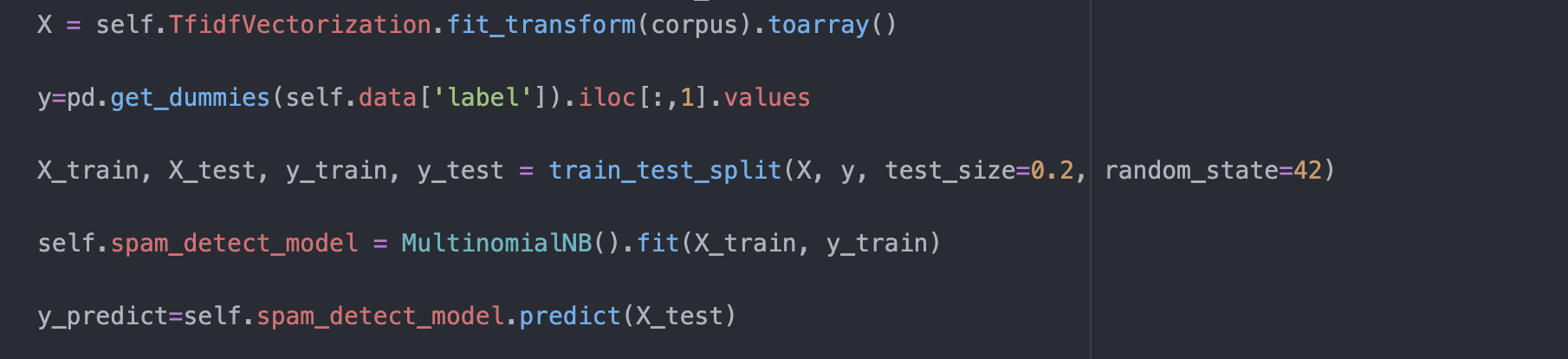


**4. Model Selection:**

- Choose the Multinomial Naive Bayes algorithm as your initial machine learning model. Import the necessary libraries from scikit-learn.

- Split your preprocessed data into a training set and a test set for model evaluation.

- Train the Multinomial Naive Bayes model on the TF-IDF-transformed training data.



**5. Evaluation:**

- Use the trained model to make predictions on the test dataset.

- Calculate various performance metrics, including:

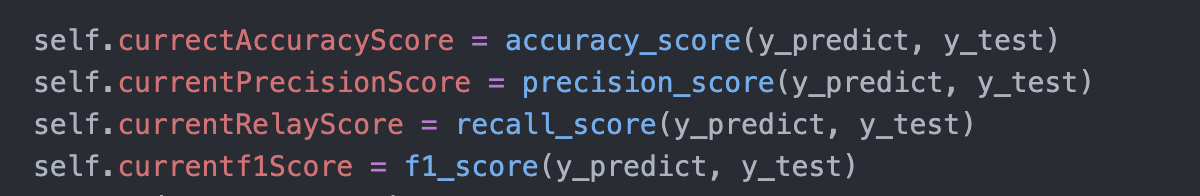
- Accuracy: The proportion of correctly classified messages.

- Precision: The proportion of true spam messages among the messages classified as spam.

- Recall: The proportion of true spam messages correctly classified as spam.

- F1-score: A harmonic mean of precision and recall, which balances both metrics.

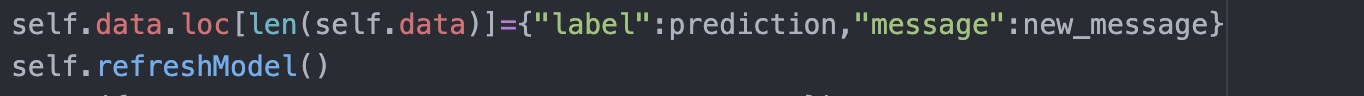
- Assess the model's performance using these metrics to determine its initial effectiveness.



**6. Iterative Improvement:**

- When a new message is predicted using your initial model, update your dataset by appending the prediction result ("spam" or "ham") along with the message.

- Periodically, or when enough new data is collected, refresh the model by repeating steps 2 to 5 using the updated dataset.



**7. Deployment:**

- Once you have achieved satisfactory performance, you can deploy your web application.

- Develop a user interface for users to input messages and receive spam classification results.

- Host the web application on a server or cloud platform.

- Ensure the model is retrained periodically with new data to maintain its accuracy and effectiveness.

