Implementation of Second Module

Machine Learning Model Building :

This module focuses on building, training, and saving the machine learning model that predicts whether a medicine is Safe or Not Safe. Along with the model, the encoder and scaler objects used during preprocessing were also saved for use in the web application.

Steps Followed:

1. **Train-Test Split:**

   - Split the dataset into 80% training and 20% testing using train_test_split.

   - Ensured class labels (Safe/Not Safe) were balanced in both sets.

2. **Model Selection:**

   - Tried multiple classification algorithms:

     - SVM

     - Random Forest Classifier

     - XBBoost

   - Evaluated models based on:

     - Accuracy Score

- o Confusion Matrix

- o Precision, Recall, F1-Score

3. **Best Model Chosen:**

- **Random Forest Classifier** gave the best results (highest accuracy and robustness).

- Trained it using the processed data (after encoding and standardization).

4. **Model Evaluation:**

- Evaluated on the test set.

- Results showed strong performance in classifying both Safe and Not Safe medicines accurately.

- Avoided overfitting by tuning model parameters where necessary.

5. **Saving the Model and Preprocessing Tools:**

- Saved the trained ML model as medicine_quality_model.pkl using joblib.

- Saved the **Label Encoder** used for Active Ingredient as encoder.pkl.

- Saved the **StandardScaler** used for numerical features as scaler.pkl.

6. Ready for Web Integration:

- These .pkl files will be loaded in the Flask backend during prediction.
- The model is now fully functional and optimized for real-time predictions.