

Methodology / Implementation

Step-by-Step Implementation Plan :

1. Data Collection:

The dataset used for this project was collected from **Kaggle**, containing 1100 entries with key features like:

- Active Ingredient
- Days Until Expiry
- Storage Temperature
- Warning Labels Present
- Dissolution Rate
- Disintegration Time
- Impurity Level
- Assay Purity
- Target: *Safe / Not Safe*

2. Exploratory Data Analysis (EDA):

We analyzed the dataset using various techniques to understand its structure:

- Checked for missing values (e.g., in Days Until Expiry)
- Plotted histograms and boxplots for feature distributions
- Used correlation matrices to study relationships between features
- Verified balance in the Safe/Not Safe class labels

3. Data Preprocessing:

- Handled missing or inconsistent data
- Encoded categorical features like Active Ingredient
- Scaled numerical features using StandardScaler or MinMaxScaler
- Split data into training and testing sets

4. Model Training:

- Trained multiple machine learning models (e.g., Decision Tree, Random Forest, Logistic Regression)
- Chose the best model based on accuracy and confusion matrix
- Saved the trained model using joblib as `medicine_quality_model.pkl`

5. Flask Web Application Development:

- Created a simple web UI using HTML and CSS
- Built Flask backend to receive inputs and return prediction
- Integrated ML model to give real-time prediction results

6. Testing and Validation:

- Validated the model using test set
- Checked predictions for known inputs
- Verified output format and error handling in the web app