1. **Title**: SafeBite: AI-Powered Allergen Detection in Food

2. **Objective**:

   To build a machine learning model that can accurately classify food products based on their allergen status using the allergen dataset.

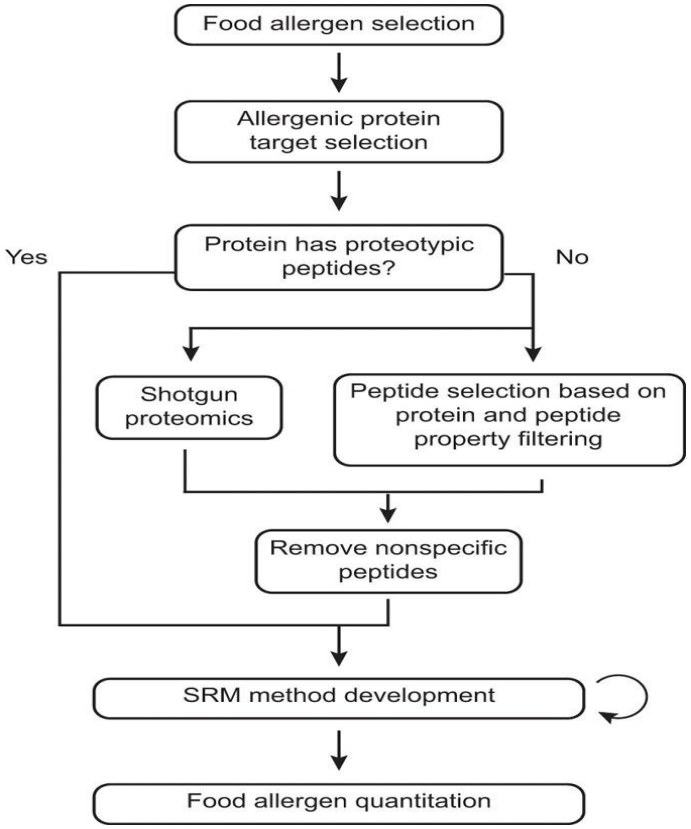
3. **Dataset**:

   Allergen Dataset: A dataset containing information about the allergen status of various food products.

4. **Project Workflow:**

   1. Data Collection and Exploration
   2. Exploratory Data Analysis (EDA)
   3. Data Preprocessing
   4. Model Building
   5. Model Evaluation
   6. Deployment and Documentation

Flow Diagram:



5. **Week-wise module implementation and high-level requirements with output screenshots**

Milestone 1: Weeks 1-2

- Data Collection and Exploration
- Understand the Problem Statement:
- Define what constitutes allergen and non-allergen food products.
- Understand the features available in the dataset.
- Collect Data:
- Download and load the allergen dataset.
- Familiarize with the structure and content of the dataset.
- Initial Data Exploration:
- Examine the dataset for initial insights.
- Identify the types of features (textual, categorical, numerical).

- Exploratory Data Analysis (EDA) I. Univariate Analysis:
  - ➢ Analyze individual features to understand their distributions.
  - ➢ Visualize the frequency of allergen vs. non-allergen food products.

II. Bivariate Analysis:
  - ➢ Explore relationships between features.
  - ➢ Identify any patterns or correlations.

III. Visualization:
  - ➢ Use plots (histograms, box plots, bar charts) to visualize data. ➢ Summarize insights from the EDA.

Milestone 2: Weeks 3-4 (Data Preprocessing)

I. Handling Missing Values:
  - ➢ Identify and treat any missing values.

II. Feature Engineering:
  - ➢ Create features such as ingredient counts, presence of specific keywords, etc.

III. Encoding Categorical Features:
  - ➢ Encode categorical features appropriately.

IV. Normalization and Standardization:
  - ➢ Normalize or standardize numerical features if necessary.

Milestone 3: Weeks 5-6 (Model Building, Model Evaluation)

I. Split Data:
  - ➢ Split the dataset into training and testing sets. II. Model Selection:
  - ➢ Explore various machine learning algorithms (e.g., Naive Bayes, Logistic Regression, Random Forest, SVM).

III. Model Training:

- ➢ Train multiple models on the training data.
- ➢ Use cross-validation to evaluate models.

IV. Hyperparameter Tuning:

- ➢ Use grid search or random search to tune hyperparameters for the best-performing models. V. Evaluate Models:
- ➢ Use metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

VI. Compare Models:

- ➢ Compare the performance of different models.
- ➢ Select the best model based on evaluation metrics.

VII. Final Model Training:

- ➢ Retrain the best model on the entire training set.

Milestone 4: Weeks 7-8 (Deployment and Documentation)

I. Model Deployment:

- ➢ Deploy the final model using a web framework like Flask or Django.
- ➢ Create an API for the allergen detection model.

II. Presentation and Documentation:

- ➢ Prepare a presentation detailing the problem statement, data collection, preprocessing steps, model building, and evaluation.
- ➢ Document the project comprehensively, including code, methodologies, and findings.

III. GitHub Submission:

- ➢ Upload the final code, documentation, and presentation to GitHub.

6. Evaluation Criteria:

Milestone 1: Week 2

- ○ Understanding of the problem statement and dataset.
- ○ Initial data exploration and insights.
- ○ Completed EDA with visualizations.

Milestone 2: Week 4

- ○ Completed data preprocessing.
- ○ Feature engineering and handling of missing values.

Milestone 3: Week 6

- ○ Model building and initial training results.
- ○ Selection of the best model based on evaluation metrics.

Milestone 4: Week 8

- ○ Deployment of the model.

- Comprehensive project documentation and final presentation.