## Project Title: Recipe Generation with LSTM Networks

### *Objective:*

Develop a model to generate cooking recipes using the RecipeNLG dataset. Students will implement an LSTM-based neural network to learn from the dataset and generate coherent and contextually relevant recipes.

### *Dataset:*

- **Name:** RecipeNLG
- **Content:** 2,231,142 cooking recipes
- **File:** RecipeNLG_dataset.csv (2.29 GB)
- **Columns:** Title, Ingredients, Directions, Link, Source, NER (Named Entity Recognition)

### *Project Outline:*

1. **Data Exploration and Preprocessing:**
    - **Load the dataset**: Use appropriate libraries (e.g., Pandas) to load and explore the CSV file.
    - **Data Cleaning**: Handle missing values, duplicate entries, and any inconsistencies.
    - **Text Processing**: Tokenize the text, remove stop words, and perform stemming/lemmatization if needed.
    - **Data Formatting**: Convert recipes into sequences suitable for LSTM input. Separate ingredients and directions for targeted training.
2. **Data Preparation:**
    - **Sequence Creation**: Create sequences of tokens from recipe directions for training the LSTM model.
    - **Train-Test Split**: Divide the dataset into training and testing sets.
3. **Model Building:**
    - **Define the LSTM Model**:
        - **Input Layer**: Embed the sequences using an embedding layer.
        - **LSTM Layers**: Implement one or more LSTM layers with dropout for regularization.
        - **Dense Layer**: Add a dense layer to output predictions.
        - **Output Layer**: Use a softmax activation function to predict the next token in the sequence.

    - **Compile the Model**: Choose an appropriate loss function (e.g., categorical cross-entropy) and optimizer (e.g., Adam).
4. **Training:**
    - **Fit the Model**: Train the LSTM model on the training data.
    - **Monitor Performance**: Track the loss and accuracy metrics during training. Use validation data to monitor overfitting.
5. **Evaluation:**
    - **Generate Recipes**: Use the trained model to generate new recipes from a seed input.
    - **Evaluate Output**: Assess the quality and coherence of the generated recipes. Consider using human evaluation or similarity metrics with existing recipes.
6. **Documentation and Reporting:**
    - **Code Documentation**: Ensure all code is well-commented and organized.
    - **Final Report**: Prepare a comprehensive report detailing the methodology, model performance, challenges faced, and results. Include code snippets, visualizations, and generated recipes.
7. **Submission:**
    - **Deliverables**: Submit the code, a detailed report, and any supplementary materials (e.g., trained models, scripts).

***Evaluation Criteria:***

- **Data Preprocessing and Cleaning**: Effectiveness in handling and preparing data.
- **Model Implementation**: Accuracy and performance of the LSTM model.
- **Recipe Generation Quality**: Coherence and relevance of generated recipes.
- **Documentation**: Clarity and thoroughness of the final report and code documentation.

***Additional Resources:***

- RecipeNLG Dataset