RECIPE RECOMMENDATION EDA ASSIGNMENT

PREPARED BY KAMAKSHI HANDA

OBJECTIVE

As an ML engineer at food.com, our task is to design a recommender system that suggests recipes to users based on their choices and the recipe they are currently viewing. A successful recommender system can increase user engagement and lead to more business opportunities. The performance of the recommendation engine will directly impact the revenue generated by the website. However, building a recommender from scratch is time-consuming. In this assignment, you will explore data and create features to build the recommender.

STEPS TO APPROACH THE PROBLEM

- Create and launch an EMR Cluster on Amazon AWS

- Create and launch a Jupyter Notebook on top of this cluster

- Perform all the necessary tasks provided in task list

TASK LIST

Task 1: Read the data

Read RAW_recipes.csv from S3 bucket.
Ensure each field has the correct data type.

Task 2: Extract individual features from the nutrition column.

Notice the nutrition column is read as a string when it should be an array of float values. Each row in the nutrition column contains seven values. Each value represents nutrition information. Our task is to separate the array into seven individual columns.

Task 3: Standardize the nutrition values.

The nutritional values in absolute terms will have a lot of variation. For example, a recipe serving six people (recipe A) will have more sugar than a recipe meant to serve one person (recipe B). But that does not necessarily imply that the sugar per person in recipe A is more than in recipe B. Standardize the nutrition values. Convert the nutritional values to per 100 calories.

Task 4: Convert the tags column from a string to an array of strings.

Task 5: Read the second data file

Read the RAW_interaction.csv and join this interaction level file with the recipe level data frame. The resulting data frame should have all the interactions.

Task 6: Create time-based features.

Create features that capture the time passed between one review and the date on which the recipe was submitted.

Task 7: Processing Numerical Columns (Optional)

Task 8: Create user-level features (Optional)

Task 9: Create tag-level features (Optional)