Recipes of AFA

Abstract

The goal of this project is to build a content-based recommender system using NLP. The recipes of AFA is a recommendation system with content-based filtering methods to generate recipes that have something in common in terms of the ingredients.

Dataset

The dataset contains around 40,000 recipes scrapped from Allrecpies.com (Kaggle). Each recipe entry contains a recipe title, a list of ingredients and measurements, instructions for preparation and a picture of the final result.

Algorithms

1- Data Preprocessing**:** Various preprocessing steps were used such as: tokenization, removing punctuation marks and digits and removing stop words that were created based on the proposed model.
2 - Vectorization : Count vectorizer and TF-IDF vectorizer
3 - Matrix factorization: LSA and NMF
4 - Cosine similarity.

Model Evaluation and Selection

The TF-IDF vectorized data with LSA gives better recommendations.

Tools

1- Numpy and Pandas for data manipulation
2- Scikit-learn for modeling
3- Matplotlib and Seaborn for plotting
4- NLTK and Gensim for topic modeling

Communication


milk
mozzarella
cheddar
parmesan
bread
onion
jar
paprika
flour
olive
oil
sesame
lemon
cheese
soy
soda
egg
pepper
cinnamon
basil
garlic
vegetable
pasta
bell
wine
salt
cayenne
vanilla
sugar
vinegar
seeded
spinach
walnuts