Flavour Fusion

TEAM MEMBERS:
IMT2020015 Pushkar Pawar
IMT2020056 Rudransh Dixit
IMT2020100 Teja Janaki Ram
IMT2020121 Manan Patel
IMT2020133 Darpan Singh

Novelty of the problem

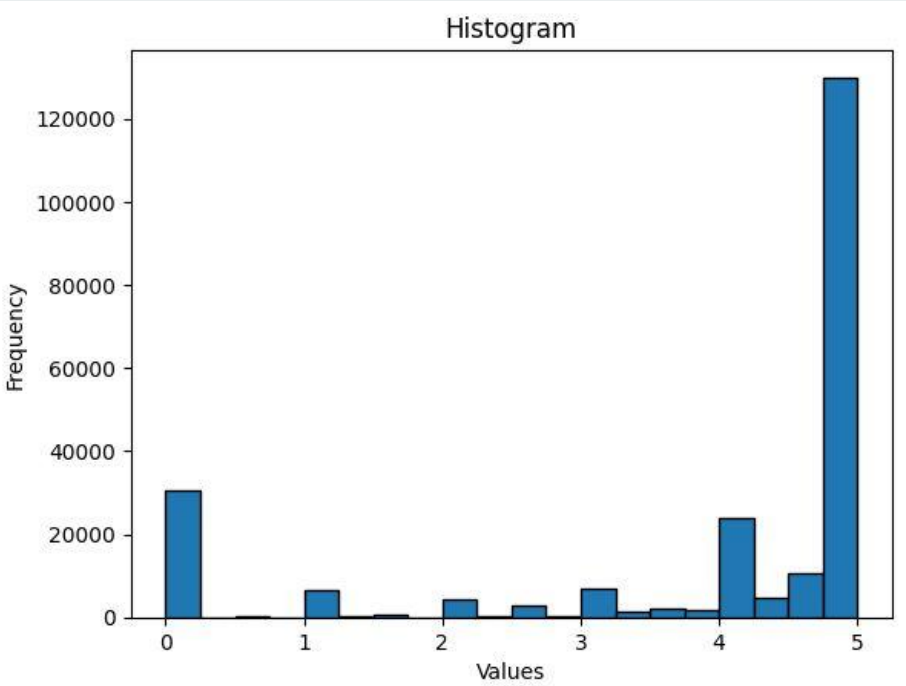
- Many people, when only a few ingredients are available in their homes, don't think it is possible to make a dish that they like or atleast any dish.
- Hence, they turn to ordering the food online
- Our goal is to recommend dishes that they like with only the ingredients available to them.
- Ultimately, we achieve making the user choose home-made food over restaurant-made food and also probably introduced him/her to new recipes and cuisines.

Dataset creation and preprocessing

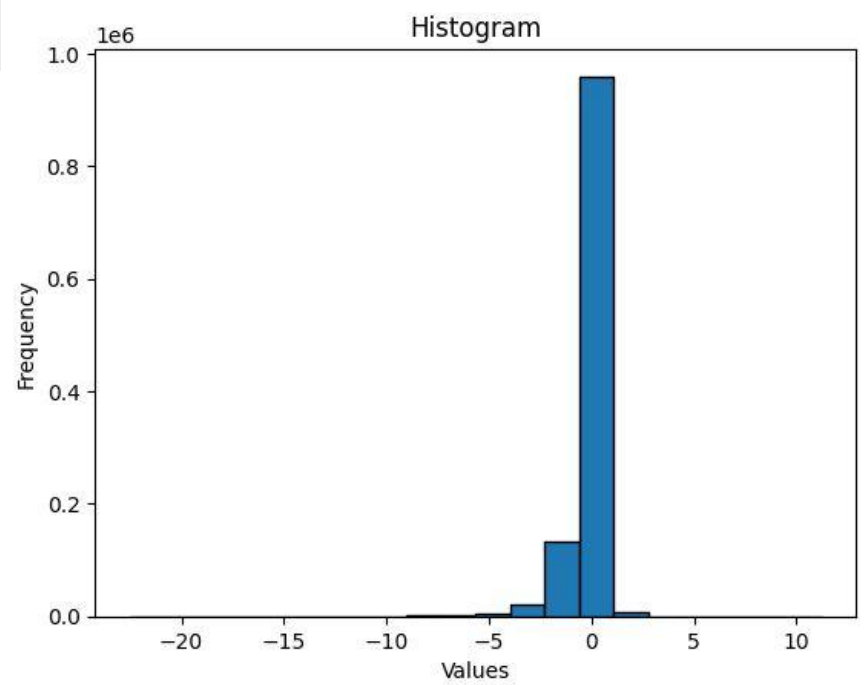
- Kaggle dataset - Crawled data from food.com (Formerly Genius Kitchen)
- 230K+ recipes and 1.1M+ ratings
- Filtered users who didn’t give enough ratings from the Top-20 most common recipes

Data Analysis

- Most of the features were in text form, therefore it does not contain any information that can be analysed statistically.
- Only rating feature can be analysed statistically.
- We have perform standard scaling of rating given by each user.

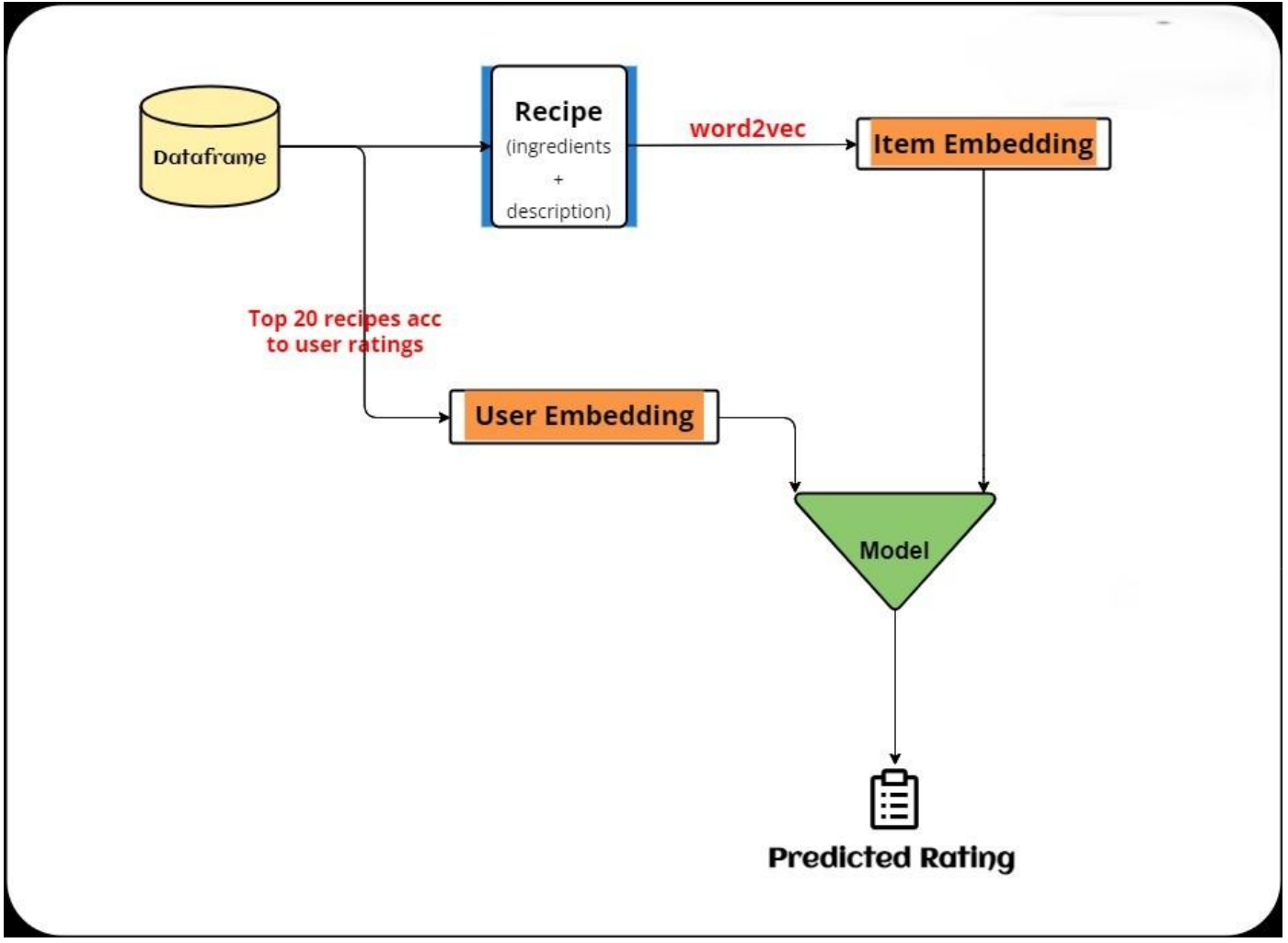


Before



After

Approach to solve the problem


Dataframe
Recipe
(ingredients
+
description)
word2vec
Item Embedding
Top 20 recipes acc
to user ratings
User Embedding
Model
Predicted Rating

User embeddings

- First we calculated top 20 most rated recipes
- Now for each user we calculated its rating for the above top 20 recipes.
- This list will act as user embedding.
- For new user we first ask them to rate this top 20 recipes to get recommendation.

Item (recipes) embeddings

- For each recipe we took its ingredient and description data
- We applied word 2 vec to get its embeddings
- Also tried BERT transformer but got poor results

Philosophy behind algorithm selection

- Due to the huge dataset we used neural networks and tree based algorithms.
- Got following results

Model	Train Accuracy	Test Accuracy
Light Gradient Boosting (LGBM)	0.9056	0.8157
Xgboost	0.8877	0.8413
Neural network	0.8673	0.8562

Model Accuracy Using BERT

Model	Train Accuracy	Test Accuracy
Neural Network	0.7410	0.6362

NN architecture

- 5 hidden layer
- Relu activation function
- Adam optimiser with MSE loss used
- Added dropout after each layer for regularization
- Gives rating as output

Algorithm Analysis

- Neural Network (Regression Net) for different number of layers

Number of hidden layers	Training accuracy	Testing accuracy
3	0.79	0.69
4	0.801	0.72
5	0.86	0.85
6	0.87	0.77

NOVELTY

Nutrition Score

- Calculating a “nutrition score” for each recipe and using it along with predicted recipes to rank the recipes.
- Weighted addition over all values of the nutrition array based on the diet choice.
- Normalizing all the values due to the huge difference in component values.

Nutrition sensitivity meter

- A nutrition sensitivity meter which can adjusted according to user's nutrition needs.
- Has the range of (0,1)
- More nutritional value ⟹ Less focus on taste

Diet choice

- Option to choose among diet types.
- Can choose among balanced diet, high protein
- Balanced diet has reasonably equal weights whereas high protein would have more weight for protein.

Time taken

- Added option for user to view only recipes based on his time requirement.
- Filter recipes which take more time than specified by the user.

THANK YOU