# Relationship Between Nutritional composition and Popularity of a Recipe

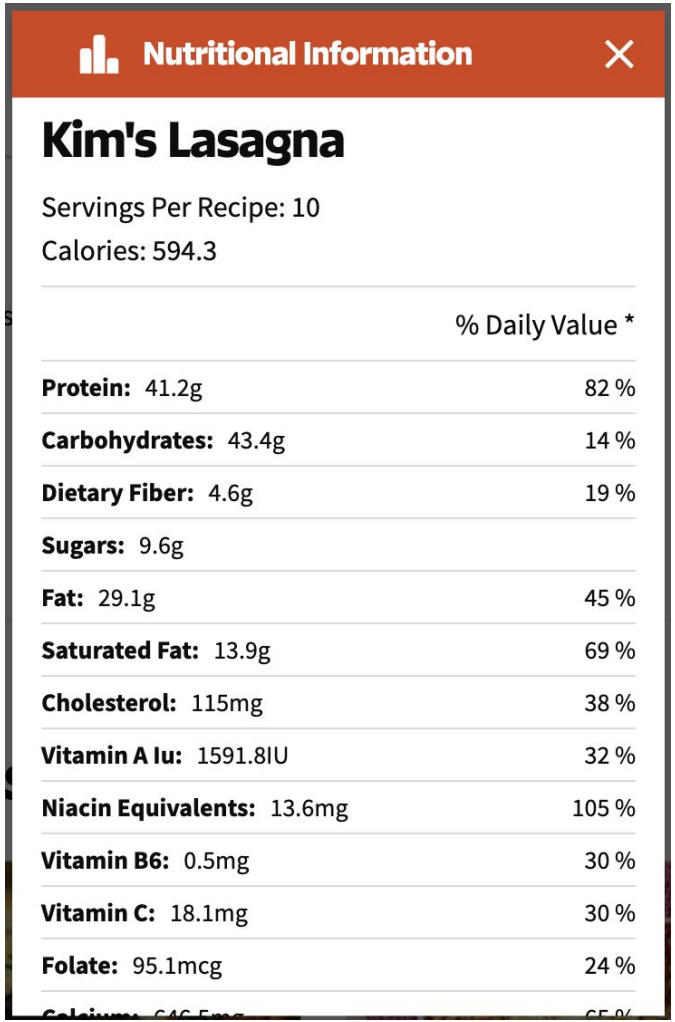
## A Linear Approach

**Milad Afrasiabi**

**September 2021**

Can nutritional composition of a recipes contribute to its popularity?

Is it possible to explain it using a linear model?



# Data

- Rating and nutritional information of 1760 dinner recipes extracted using Beautiful soup
- 260 were discarded due to website format irregularities
- Outliers were removed using EDA
- 1472 rows of data were selected for final analysis, 18 original features

Target:

Rating

## Original Features:

- Protein
- Carbs
- Fibers
- Sugars
- Fat
- Saturated Fat
- Cholesterol
- Vitamin A
- Niacin
- Vitamin B6
- Vitamin C
- Folate
- Calcium
- Iron
- Magnesium
- Potassium
- Sodium
- Thiamin

# Methodology

1800 recipes scrapped, 300 discarded as irregular, outliers removed

Data divided 80/20, train vs. holdout (scikit learn model selection package)

Feature engineering and addition
30 in total

Train data divided to 5 CV groups (using scikit learn KFold package)

OLS, Ridge, LASSO and Elastic Net models fitted and scores averaged

Best models were selected and tested on the holdout set

# Model Selection

OLS-$R^2$ = 0.066 ± 0.048

LASSO-R2 = 0.063 ± 0.041

Ridge-R2 = 0.067 ± 0.043

ElasticN-$R^2$ = 0.066 ± 0.042

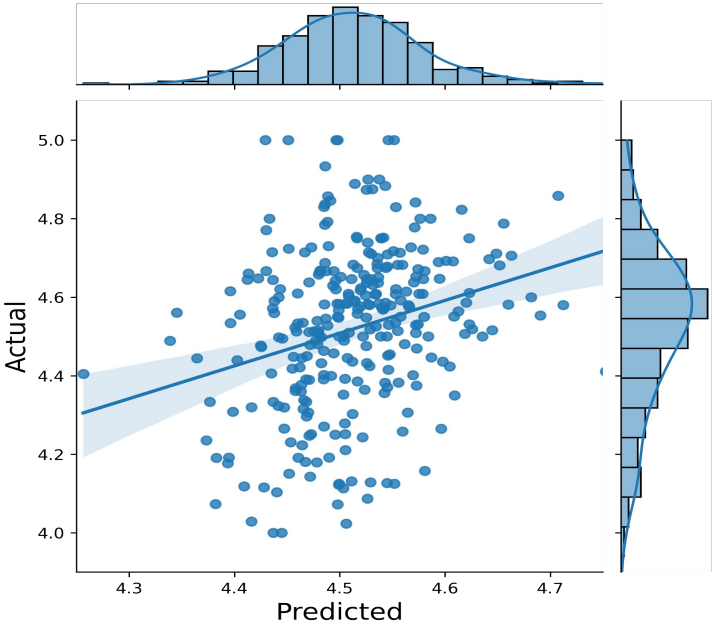
# Results

Average on 5 validation sets:

$R^2 = 0.059 \pm 0.032$

Trained on whole data, scored on test:

$R^2 = 0.059$
MAE = 0.162




## Key Features:

- **Protein**
- **Fat**
- **Cholesterol**

Can nutritional composition of a recipes contribute to its popularity? *Maybe!*

Is it possible to explain it using a linear model? *Hmm...*

- **Adding non-nutritional features**
- **Using non-linear models (Decision Trees, Random Forest, etc.)**