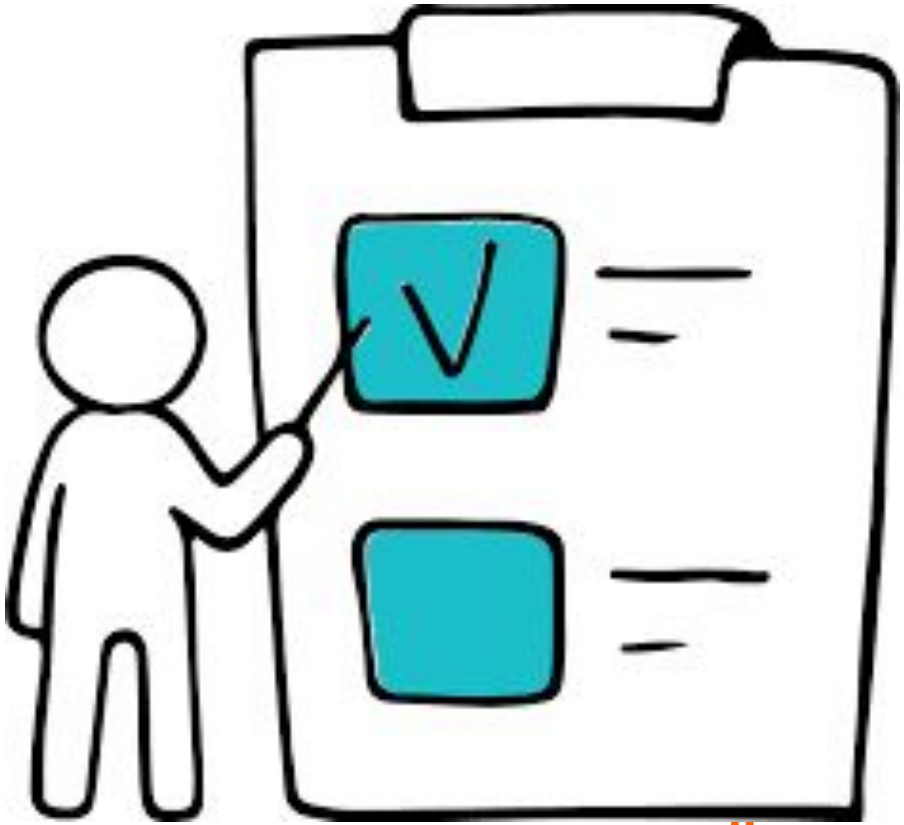
Recipes Rating System

Team 4
Liyuan Li, Siying Liu, Yiting Fu

allrecipes

Agenda

1. Assumptions & Methodology
2. Data Clean & Preparation
3. Modeling: KNN
4. Results



Goal

Predict a User's rating on a recipe

user_id + recipe_id → ?rating?

allrecipes

Customer-Oriented Methodology

Everyone is Unique !

Each user has a customized feature matrix of recipes

Each user should have a customized model

allrecipes

K-Nearest Neighbor Classifier

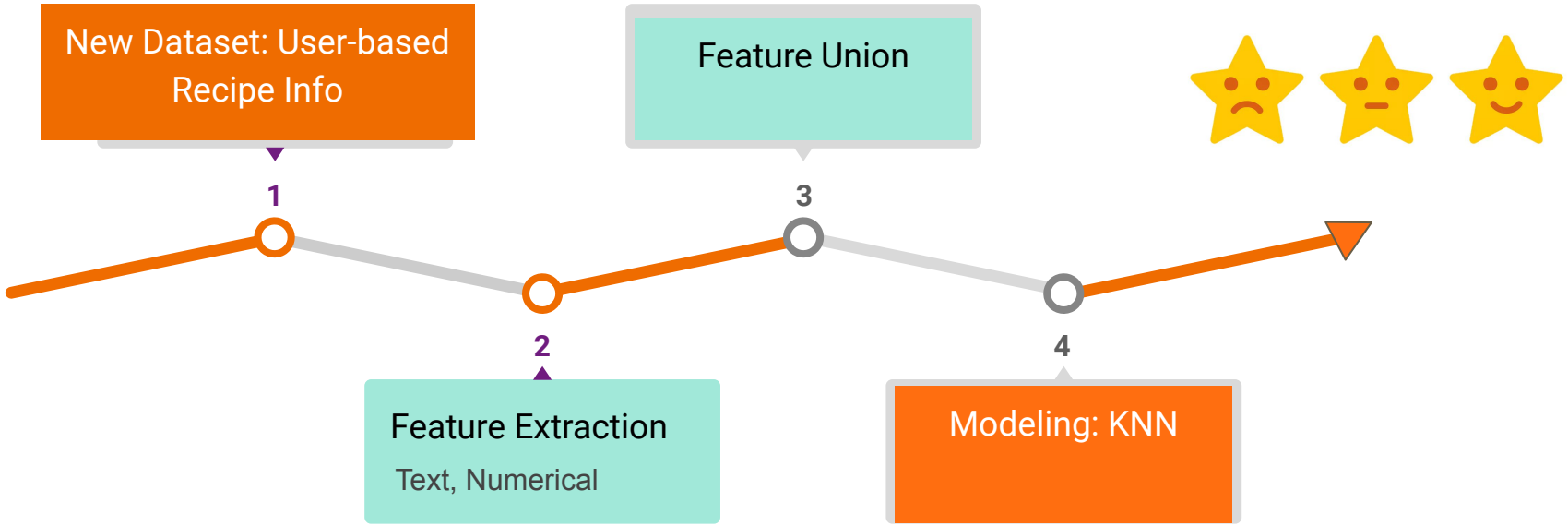
Everyone has a preference!

Features of recipe information disclose similarities and differences between recipes.

A user tend to give a similar rating for recipes having similar features.

allrecipes

Workflow



*** Our model is updating as we apply it on different users at different time...**

allrecipes

Cleaned Dataset

	recipe_id	recipe_name	aver_rate	review_nums	cooking_directions	Ingredients	Prep	Cook	Ready	Calories	Calories_Fat	Carbohydrates	Protein	user_id	rating
0	218939	Foolproof Rosemary Chicken Wings	4.571429	12	{'directions': u"Prep\n20 m\nCook\n40 m\nReady...	[chicken wings, sprigs rosemary, head garlic, ...	20	40	40	335	212	6	23	2783111	5
1	87211	Chicken Pesto Paninis	4.625000	163	{'directions': u'Prep\n15 m\nCook\n5 m\nReady ...	[focaccia bread quartered, prepared basil pest...	15	5	5	640	264	60	32	742713	4
2	87211	Chicken Pesto Paninis	4.625000	163	{'directions': u'Prep\n15 m\nCook\n5 m\nReady ...	[focaccia bread quartered, prepared basil pest...	15	5	5	640	264	60	32	674114	5
3	87211	Chicken	4.625000	163	{'directions': u'Prep\n15	[focaccia bread quartered	15	5	5	640	264	60	32	191706	4

Features Extraction

Numerical:

- Review_nums
- Aver_rating
- Time (Prep, Cook, Ready)
- Nutritions (Calories, Calories_fat, Carbohydrates, Protein)

Text:

Recipe_name

Ingredients

Cooking_directions

Vectorizer

Numerical: DictVectorizer

Lists of mappings (dict-like objects) of feature names/value

Numpy arrays/scipy.sparse matrices for use with scikit-learn estimators

Text: TfidfVectorizer

a collection of raw documents

a matrix of TF-IDF features

Feature Union

Concatenates results of multiple transformer objects

Combine several feature extraction mechanisms into a single transformer

class `sklearn.pipeline.` **`FeatureUnion`** (*transformer_list*, *n_jobs=None*, *transformer_weights=None*, *verbose=False*) [source]

allrecipes

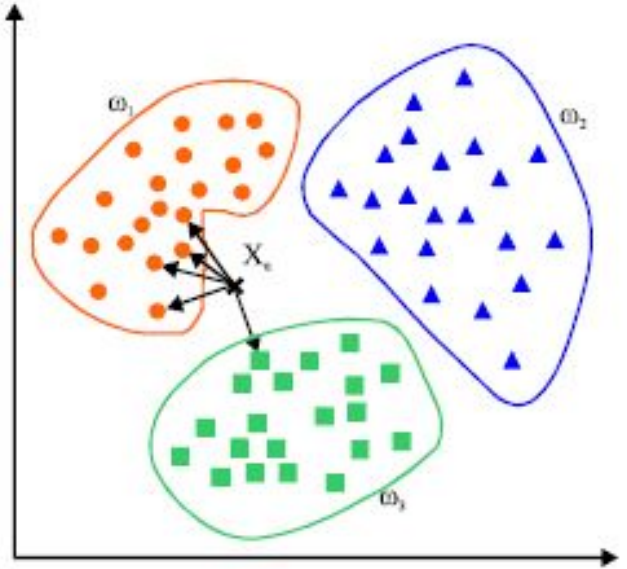
KNN

K = ?

Users have many/few ratings -- different K

Grid Search:

Find the optimal parameters of each model for individual user



allrecipes

Future Improvement

Try smaller range of K when use Grid Search

Find the link between different users and then use association to predict their ratings

allrecipes

Result:

MAD = 0.373

Thank You!

Special Thanks to Wai & His Team!

allrecipes