

Topic Modeling on Food Recipes

Ed Lee (DSML), Aug 2022

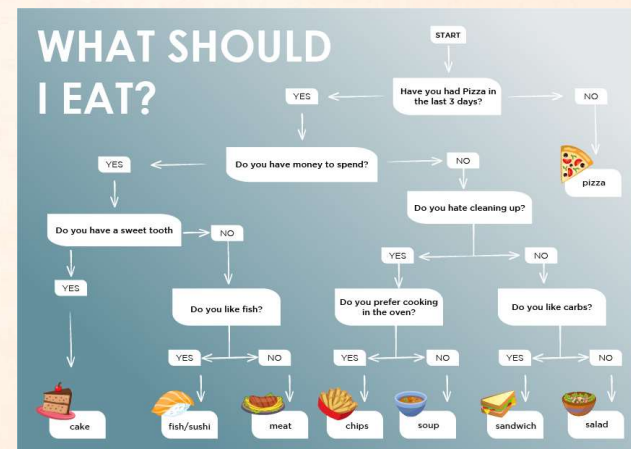


Intro – Motivation



Sometimes it is hard to decide what to eat/cook:

- To get some rough idea for a starter...
 - a. Think of some vague ingredients
 - b. Or maybe a broad category of food
- Potentially suggest some foods and their recipes from those rough idea



<https://blogs.salford.ac.uk/made-in-salford/2017/06/01/what-to-eat/>

Objective



Obtain well-defined topics from various recipe articles for further utilization



<https://blogs.salford.ac.uk/made-in-salford/2017/06/01/what-to-eat/>

Dataset

- Dataset contains recipes and ingredients from *food.com*, obtained via kaggle.
- ~ 520k recipe articles (rows), of which total words count add up to about 53,600,000.



Work Flow

01.

Data Cleaning

Removing undesirable terms

03.

Topic Modeling

Comparing NMF & LDA

02.

Vectorization

Turning the corpus into a TF-IDF matrix

04.

EDA & Model Evaluation

Assigning topics to recipes and inspection

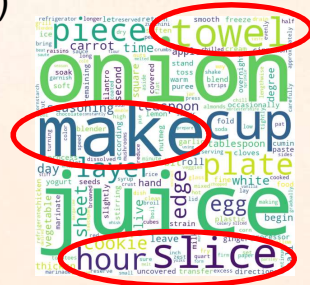


1. Data Cleaning



Term Removal

- Punctuations
- Symbols
- Numbers
- Some words commonly used in recipes, yet insignificant (ex: units)



Stemming

- Effect was manually evaluated by running models with or without conducting stemming
- Did not help nor hurt; ultimately left out



2. Vectorization



Vectorizer: TF-IDF

- Chosen over CountVectorizer
- Takes key ingredient into account



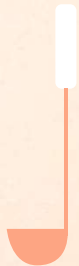
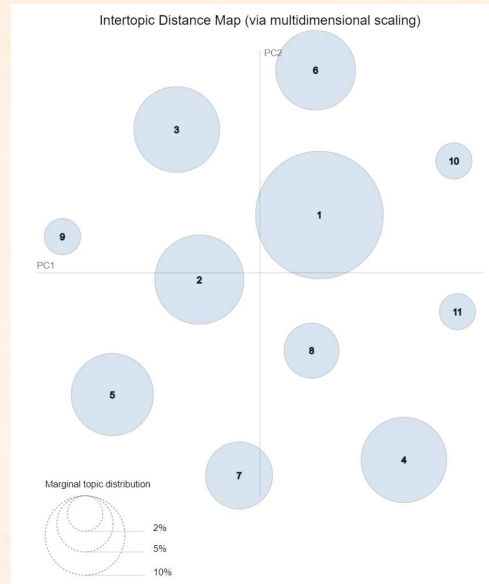
Tuning TF-IDF

- Hyperparameters (df_max/df_min) were tuned by manual inspection
 - Modeling and then inspecting topics' terms



3. Topic Model – LDA

- Topic clusters show good separation



3. Topic Model – LDA

- However, some topics were difficult to interpret

				???						???	???	???
0	batter	blender	cheese	chicken	chicken	chocolate	pork	dough	tortilla	shake	potatoes	
1	beat	ice	bread	grill	simmer	cream	beans	roll	salsa	enjoy	fish	
2	flour	glass	dish	dressing	sauce	pie	hours	sheet	chips	jar	taste	
3	cake	blend	sprinkle	marinade	boil	crust	soak	cookie	chill	store	cut	
4	eggs	juice	degrees	salad	pot	cool	overnight	balls	peanut	container	garlic	
5	vanilla	smooth	spread	toss	garlic	milk	rinse	flour	lettuce	lid	olive	
6	cool	apples	bacon	fridge	onion	syrup	fat	shape	fold	use	juice	
7	muffin	processor	casserole	bag	skillet	saucepan	approximately	inch	extra	tightly	lemon	
8	dry	food	spray	hours	rice	vanilla	inches	cut	square	color	use	
9	egg	pineapple	slices	refrigerate	bring	stirring	board	ball	toothpick	teaspoons	chop	

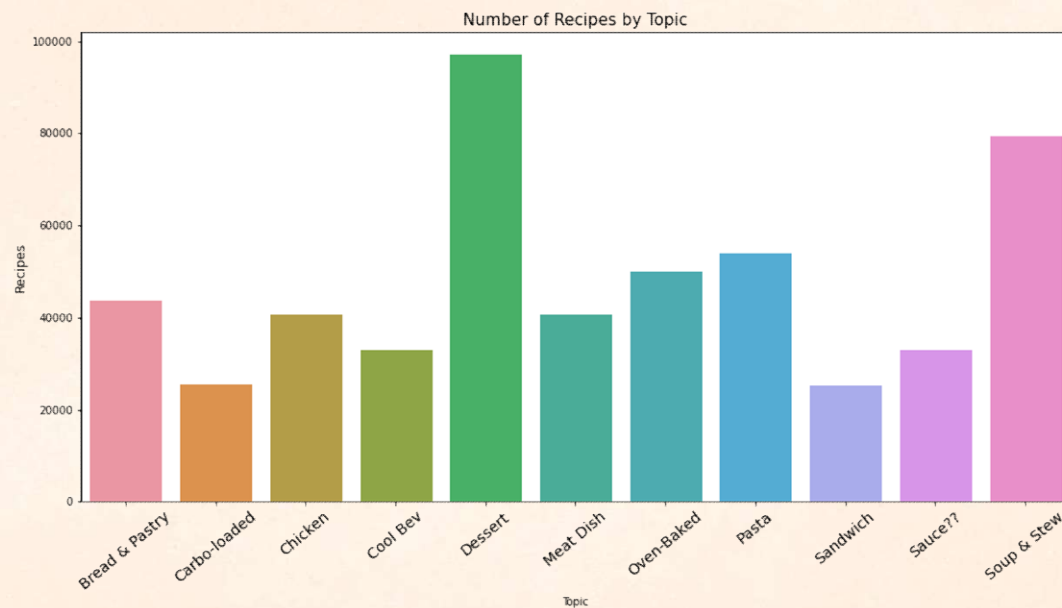
3. Topic Model – NMF

- 11 Topics
- Each topic is relatively well-characterized

	Term Rank	Meat Dish	Dessert	Chicken	Oven-Baked	Bread & Pastry	Cool Bev	Soup & Stew	Carbo-loaded	Pasta	Sauce??	Sandwich
0	0	sauce	beat	chicken	cheese	dough	ice	simmer	potatoes	pasta	blender	bread
1	1	meat	cake	pieces	dish	roll	glass	boil	potato	toss	blend	crumbs
2	2	grill	flour	cooked	cream	sheet	juice	stirring	tender	dressing	smooth	loaf
3	3	beef	cool	broth	sprinkle	cookie	lemon	pot	bacon	salad	processor	spread
4	4	pork	batter	marinade	casserole	flour	garnish	rice	mash	drain	food	cut
5	5	skil	chocolate	rice	spread	cut	shake	reduce	peel	serving	process	toast
6	6	soy	vanilla	skil	melted	inch	orange	onion	cut	olive	puree	egg
7	7	marinade	eggs	coat	layer	lightly	lime	tender	sweet	vinegar	enjoy	eggs
8	8	hot	egg	dish	sour	floured	cream	saucepan	drain	whisk	refrigerate	golden
9	9	garlic	cream	bag	degrees	surface	strain	low	dish	tomatoes	container	half

4. EDA & Model Evaluations (NMF)

❖ Frequency Distribution of Topics



People love desserts...



4. EDA & Model Evaluations (NMF)

❖ Topics Representation of Documents

Top topic for Meat Dish: Cantrell Sauce (steak Marinade and Grilling Sauce)
Model score: 0.05509799676480944



Top topic for Dessert: Decadent Devil's Food Cake
Model score: 0.07679308730243227



Top topic for Chicken: Refreshing Chicken Salad
Model score: 0.11007161101868407



Top topic for Oven-Baked: Parmesan Roasted Asparagus Spears
Model score: 0.08656220705002918



Top topic for Bread & Pastry: Foolproof Flaky Butter Pastry
Model score: 0.08925161738410683



Top topic for Cool Bev: Planters Punch
Model score: 0.09892234305283633



Top topic for Soup & Stew: Wild Mushroom Soup
Model score: 0.0693692760552689



Top topic for Carbo-loaded: Stacey's Cheesy Potatoes
Model score: 0.11485584208821759



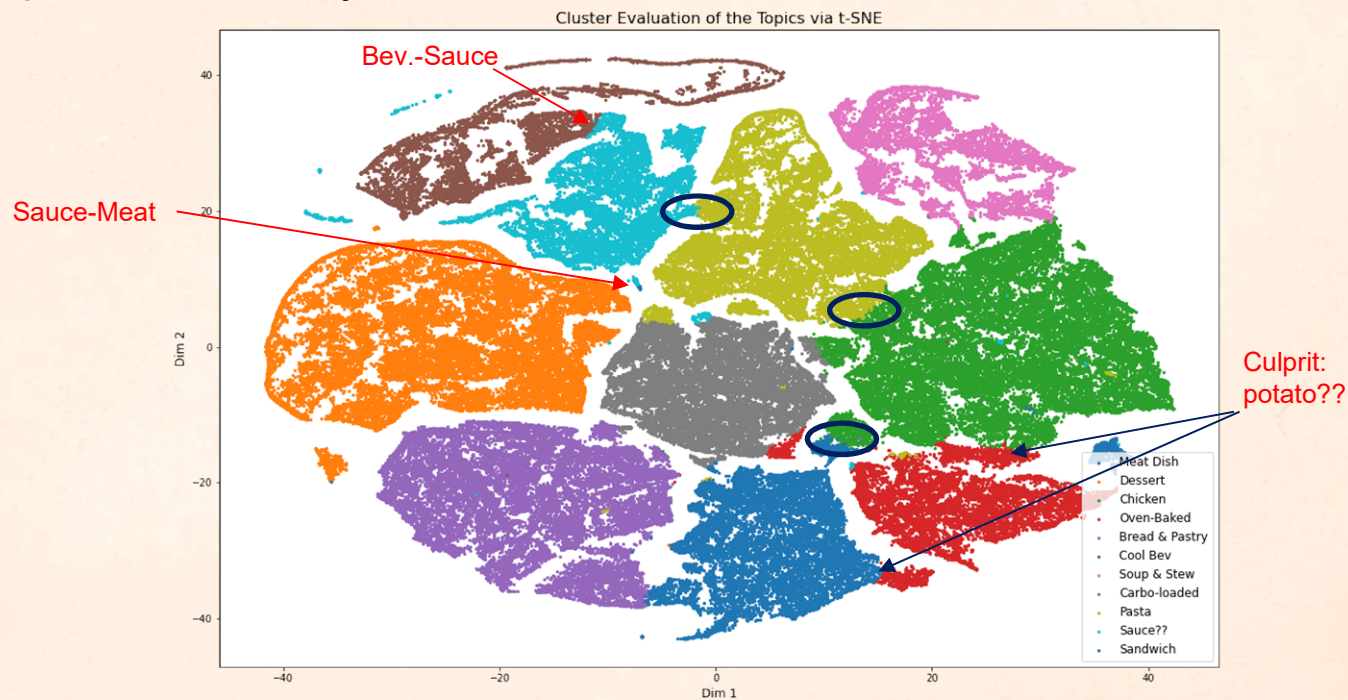
Top topic for Pasta: Insalata Florentine Romano's Macaroni Grill Copycat
Model score: 0.10880563930399409

Top topic for Sauce??: Orange-Pineapple Coconut Smoothie
Model score: 0.11575467323265687

Top topic for Sandwich: Babka Drozdowa for the Bread Machine
Model score: 0.12643289031591665

4. EDA & Model Evaluations (NMF)

❖ Topic Cluster Analysis



Conclusion & Future Works



✓ Modeling Result

- NMF model yielded a reasonably well-defined topics
- Topic represents the documents to some degree; further tuning may be possible

✓ Future Works

- Further topic model tuning (hyperparameters, term selection, etc)
- Building a **recommender**
 - > With a given article, ingredient list, or food description, suggest recipes that are similar to the input



Thanks!

Questions/Comments?



CREDITS: Diese Präsentationsvorlage wurde von Slidesgo erstellt,
inklusive Icons von Flaticon und Infografiken & Bilder von Freepik

