No Free Lunch

Leon Chen, Andrew Reuben,
Daniel Lai, Samruddhi Somani

Agenda

- Context
- Exploratory Data Analysis
- Design Choices
- Models
- Takeaways

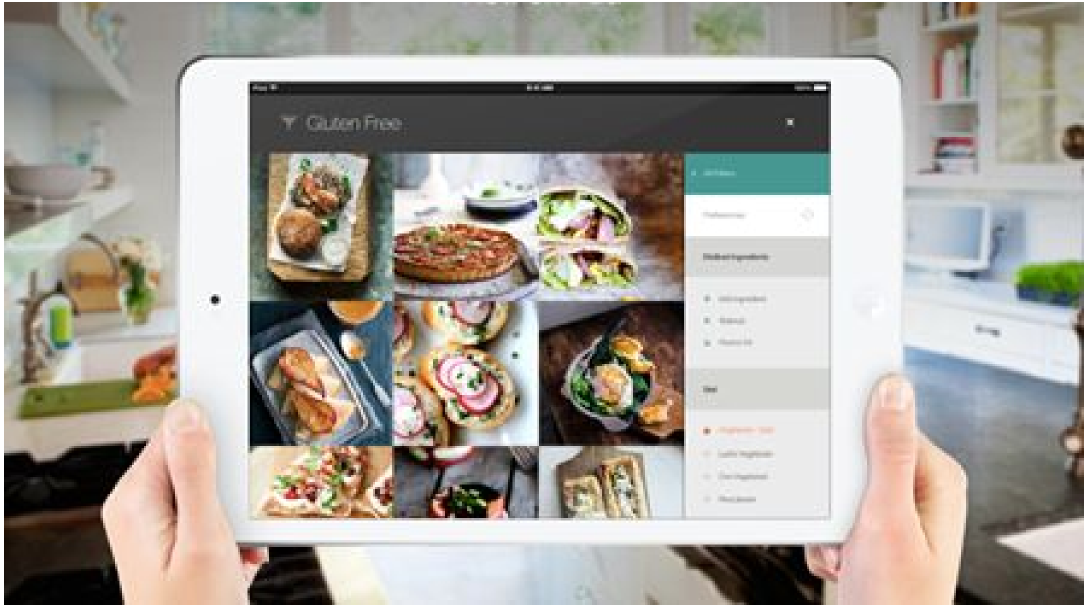
Context

Background

- Yummly is a website that:
  - Asks for your cuisine preferences
  - Gathers recipes across the web
  - Presents you with curated recommendations
- Similar to Pinterest, but targeted especially towards food!

Question

- Given just the ingredients, can cuisines be predicted?



Data

39,744 recipes in training set

9,944 in test set

Average of 7 ingredients per recipe

6,714 unique ingredients

Samples

TRAIN

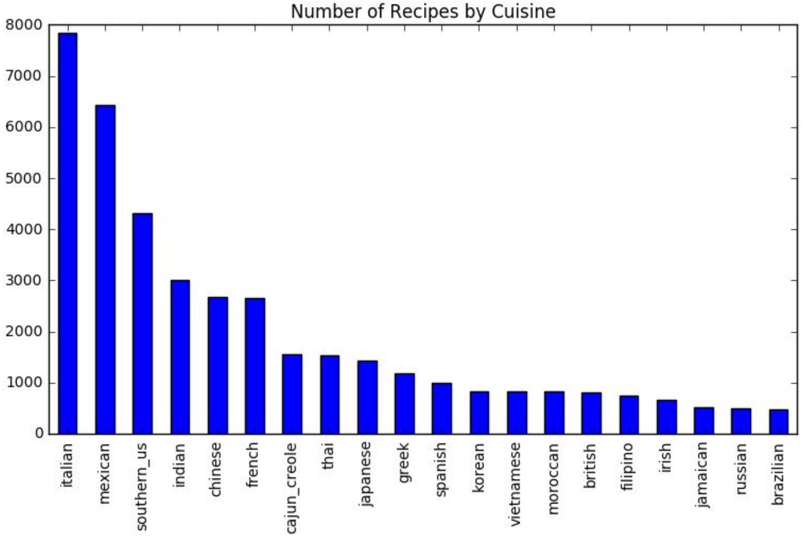
```
{
  "id": 10259,
  "cuisine": "greek",
  "ingredients": [
    "romaine lettuce",
    "black olives",
    "grape tomatoes",
    "garlic",
    "pepper",
    "purple onion",
    "seasoning",
    "garbanzo beans",
    "feta cheese crumbles"
  ]
 }
```

TEST

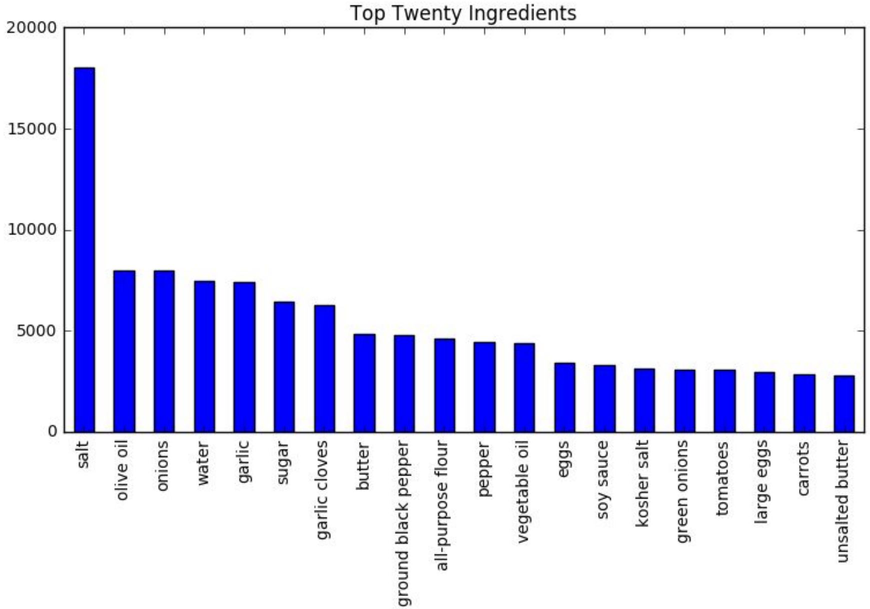
```
{
  "id": 18009,
  "ingredients": [
    "baking powder",
    "eggs",
    "all-purpose flour",
    "raisins",
    "milk",
    "white sugar"
  ]
 }
```

Exploratory Data Analysis

Aggregates



Number of Recipes by Cuisine
8000
7000
6000
5000
4000
3000
2000
1000
0
italian
mexican
southern_us
indian
chinese
french
cajun_creole
thai
japanese
greek
spanish
korean
vietnamese
moroccan
british
filipino
irish
jamaican
russian
brazilian



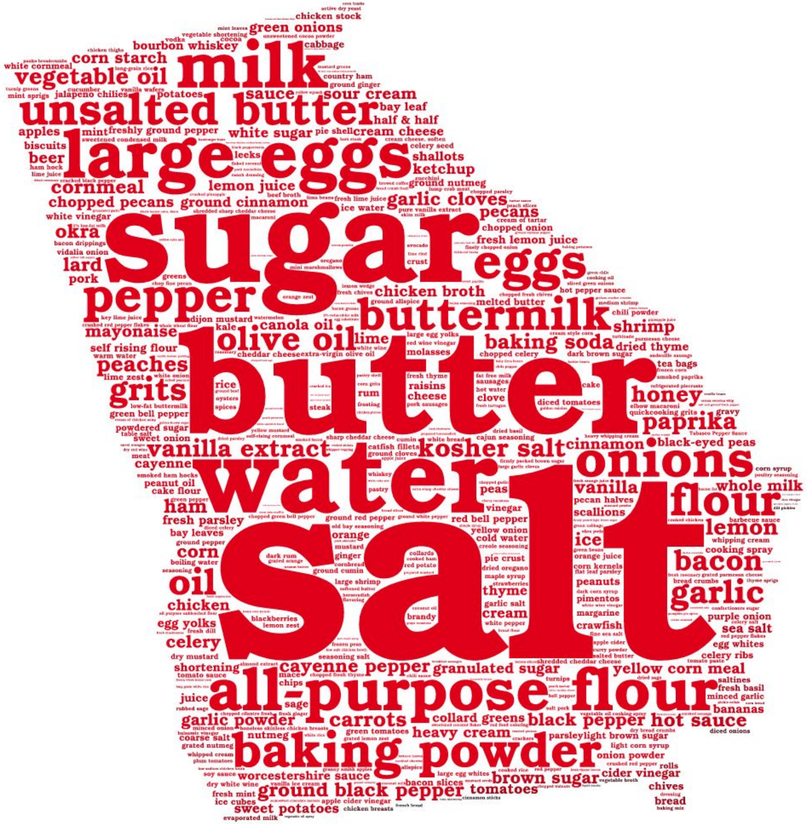
Top Twenty Ingredients
20000
15000
10000
5000
0
salt
olive oil
onions
water
garlic
sugar
garlic cloves
butter
ground black pepper
all-purpose flour
pepper
vegetable oil
eggs
soy sauce
kosher salt
green onions
tomatoes
large eggs
carrots
unsalted butter

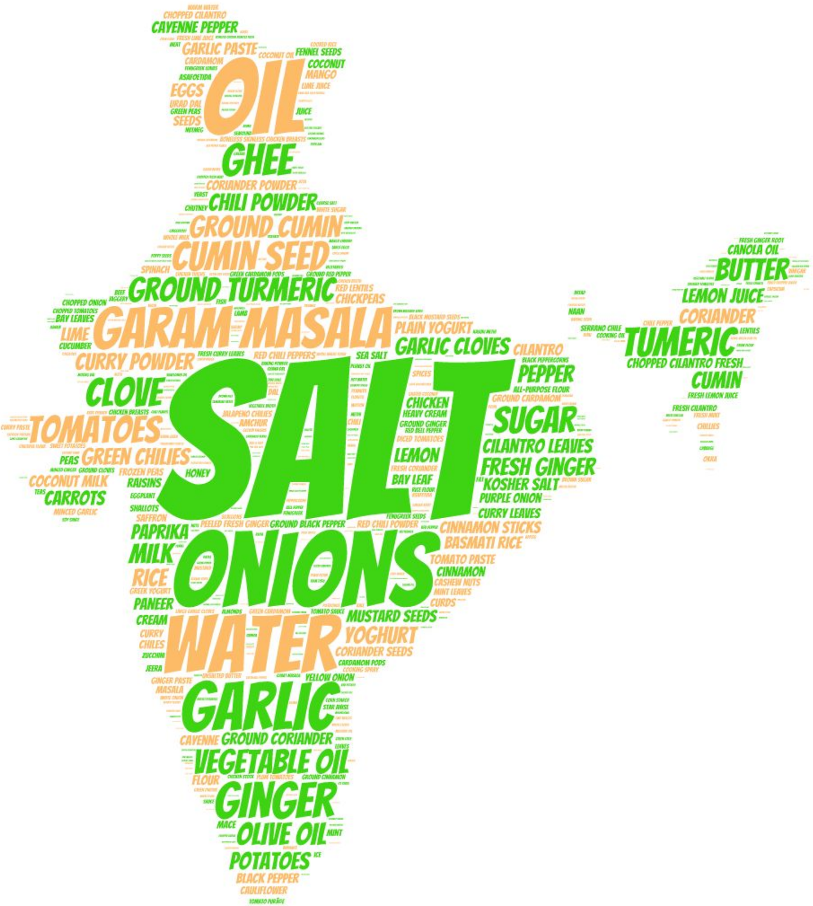
Cuisines


salt
olive oil
pepper
garlic
water
sugar
butter
onions
eggs
garlic cloves
fresh basil
kosher salt
extra-virgin olive oil
ground black pepper
grated parmesan cheese
fresh parsley
large eggs
all-purpose flour
tomatoes
milk
lemon
carrots
capers
spaghetti
linguine
parmesan cheese
dry white wine
flour
pasta
basil
zucchini
shallots
black pepper
cooking spray
mozzarella cheese
unsalted butter
dried oregano
pesto
honey
leeks
ricotta cheese
fresh lemon juice
flat leaf parsley
heavy cream
white sugar


salt
onions
garlic
olive oil
salsa
sour cream
water
cumin
lime
avocado
pepper
sugar
cilantro
chili powder
ground cumin
jalapeno chilies
corn tortillas
flour tortillas
eggs
milk
rice
flour
chicken
butter
tomatoes
lard
oil
corn
kosher salt
chili
diced tomatoes
black beans
lime juice
garlic cloves
cheese
cotija
chopped cilantro
enchilada sauce
dried oregano
vegetable oil
purple onion
ground beef
paprika
chiles
shredded cheddar cheese
cayenne pepper
chicken breasts
chicken broth
fresh cilantro
green onions
canola oil
tomatillos
sea salt

Cuisines


salt
butter
sugar
water
large eggs
milk
unsalted butter
all-purpose flour
baking powder
onions
flour
eggs
buttermilk
pepper
olive oil
vanilla extract
kosher salt
garlic
bacon
oil
lemon
ground black pepper
vegetable oil
brown sugar
carrots
paprika
honey
grits
peaches
cornmeal
garlic cloves
green onions
black pepper
hot sauce
heavy cream
baking soda
chicken broth
cayenne pepper
granulated sugar
yellow corn meal
corn starch
sour cream
cream cheese
okra
lard
pork
shrimp
vanilla
ham
corn
chicken
celery
pecans
ketchup
shallots

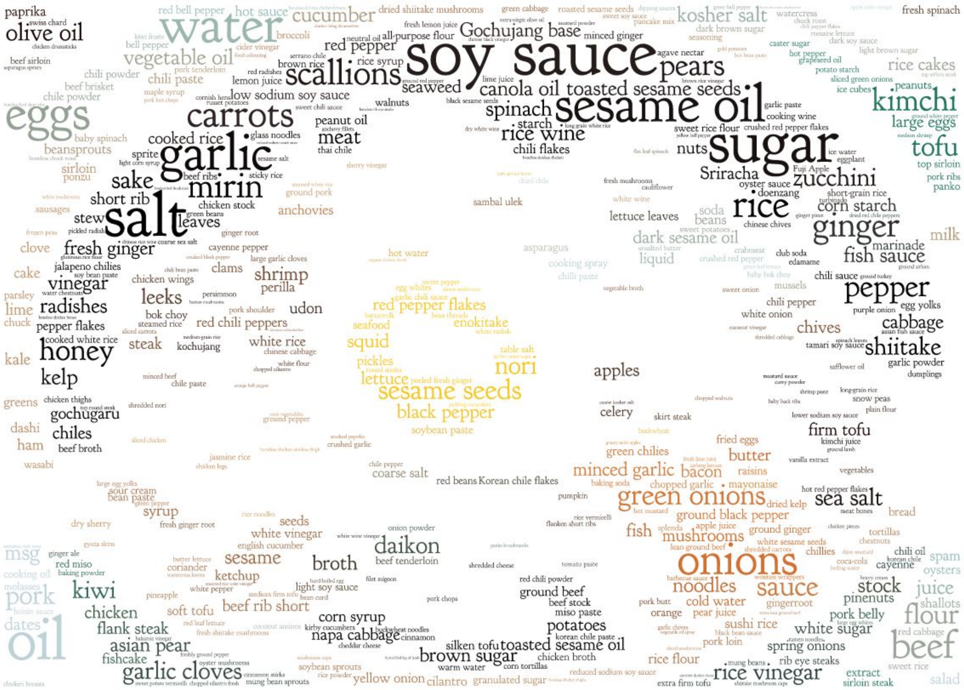

SALT
ONIONS
WATER
GARLIC
OIL
GARAM MASALA
GINGER
VEGETABLE OIL
GHEE
CUMIN SEED
GROUND CUMIN
GROUND TURMERIC
CHILI POWDER
CLOVE
TOMATOES
OLIVE OIL
POTATOES
TUMERIC
BUTTER
SUGAR
FRESH GINGER
CILANTRO LEAVES
PEPPER
LEMON
GREEN CHILIES
CURRY POWDER
MILK
PAPRIKA
RICE
PANEER
YOGHURT
MUSTARD SEEDS
GROUND CORIANDER
CUMIN
CORIANDER
LEMON JUICE
GARLIC CLOVES
PLAIN YOGURT
CINNAMON STICKS
BASMATI RICE
KOSHER SALT
COCONUT MILK
CARROTS
CAYENNE PEPPER
GARLIC PASTE
EGGS

Cuisines


soy sauce
salt
water
sugar
sesame oil
garlic
oil
eggs
corn starch
ginger
scallions
green onions
honey
carrots
onions
rice vinegar
garlic cloves
hoisin sauce
fresh ginger
vegetable oil
pepper
flour
oyster sauce
peanut oil
olive oil
canola oil
Shaoxing wine
chicken broth
light soy sauce
dark soy sauce
shiitake
chicken stock
kosher salt
white pepper
brown sugar
shrimp
rice wine
pork
Sriracha
broccoli
ketchup
noodles
star anise
chicken breasts
large eggs
red bell pepper
dry sherry
white sugar
ground pork
cold water
snow peas
fish sauce
minced garlic
sesame seeds
black pepper
shallots
sherry
rice
leeks
flank steak
tofu
chicken
milk
celery
beef
mushrooms
toasted sesame oil
chinese five-spice powder
spring onions
boneless skinless chicken breasts
low sodium soy sauce
dark sesame oil
ground ginger
cornflour
napa cabbage
sea salt
beansprouts
water chestnuts
cilantro
peanuts
lard
duck
lime
msg
wine
orange
ground black pepper
all-purpose flour
red pepper flakes
canola
cabbage
bok choy
vinegar
chili oil
juice
greens


soy sauce
sugar
mirin
sake
salt
water
ginger
oil
rice vinegar
vegetable oil
onions
scallions
carrots
nori
garlic
dashi
konbu
eggs
milk
miso
flour
gari
canola oil
honey
sushi rice
tofu
rice
sesame oil
sesame seeds
cucumber
beef
fresh ginger
butter
white miso
green onions
pepper
lemon
kosher salt
corn starch
shallots
shiitake
olive oil
avocado
seaweed
potatoes
shiso
wasabi
udon
daikon
peas
soba
pork
spinach
sauce
mayonnaise
carrots
chicken
cream
beer
chili oil
noodles
shrimp
corn
bonito
salmon
sea salt
fish
sweet potato
cabbage
panko
kale
leeks
soup
okra
chives
ground black pepper
salmon fillets
toasted sesame oil
cumin seed
Sriracha
zucchini
miso paste
cilantro
bonito flakes
lime
tahini
kelp
apples
ghee

Cuisines

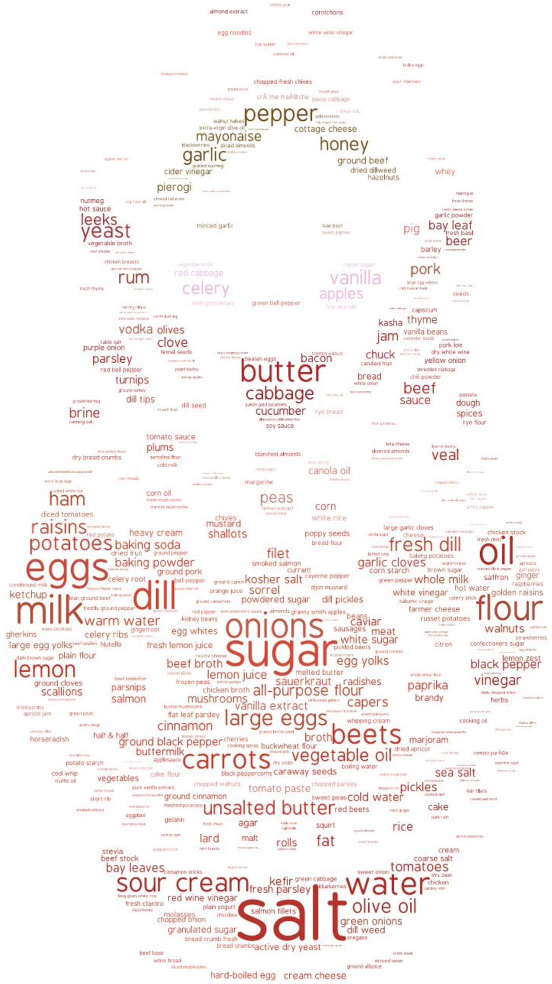


soy sauce
sesame oil
sugar
garlic
salt
water
eggs
scallions
carrots
mirin
rice
ginger
oil
onions
green onions
kimchi
tofu
pepper
honey
kelp
vinegar
radishes
sesame seeds
black pepper
nori
daikon
beef
flour
pork
sauce
rice vinegar
brown sugar
garlic cloves
napa cabbage
toasted sesame oil
zucchini
pears
canola oil
toasted sesame seeds
rice wine
fish sauce
shiitake
cabbage
sea salt
noodles
broth
kiwi

fish sauce
garlic
coconut milk
sugar
lime
oil
water
salt
soy sauce
eggs
onions
vegetable oil
cilantro
lime juice
garlic cloves
shallots
carrots
honey
shrimp
ginger
green onions
peanuts
basil
thai basil
sesame oil
olive oil
scallions
lemongrass
fresh lime juice
brown sugar
pepper
rice vinegar
zucchini
tofu

Cuisines


salt
olive oil
water
dried oregano
pepper
lemon
garlic
onions
butter
garlic cloves
dill
cucumber
feta cheese crumbles
feta cheese
fresh lemon juice
tomatoes
eggs
sugar
extra-virgin olive oil
red wine vinegar
ground black pepper
lemon juice
purple onion
greek yogurt
milk
honey
zucchini
mint
kosher salt
fresh dill
kalamata
orzo
eggplant
fresh parsley
large eggs
capers
shallots
oil
tahini
oregano
unsalted butter
couscous
all-purpose flour
black pepper
chickpeas
fresh oregano
carrots
cheese
chicken
pitas
lemon zest
pita bread
fresh mint
spinach
sea salt
cinnamon
ouzo


salt
sugar
onions
butter
eggs
milk
dill
flour
oil
water
sour cream
carrots
beets
large eggs
lemon
cabbage
pepper
garlic
mayonaise
honey
yeast
leeks
rum
celery
vanilla
apples
pork
beef
ham
raisins
potatoes
fresh dill
vegetable oil
unsalted butter
olive oil
all-purpose flour
vinegar
black pepper
walnuts
ground black pepper
baking soda
baking powder
garlic cloves
kosher salt
peas
veal
brine
tomatoes
pickles
cucumber
mushrooms
vanilla extract
cinnamon

Design Choices

Bag of “Words”

- List of ingredients unstructured
  - order of ingredients not important
  - no grammatical information lost
  - meaning of one ingredient not affected by adjacent ingredient

Tokenization

- **Keep original tokens**
  - **Best performance**
- Tokenize by word
  - no improvement over original tokenization
- Lemmatize each original token
  - poor performance in initial testing

```
[u'romaine lettuce',
 u'black olives',
 u'grape tomatoes',
 u'garlic',
 u'pepper',
 u'purple onion',
 u'seasoning',
 u'garbanzo beans',
 u'feta cheese crumbles']
```

Document Term Matrix

- Scaling
  - **Original Counts**
  - TF-IDF Scaling
- Minimum Documents: 5 to filter spelling errors and uncommon ingredients

Models

Establishing a Baseline

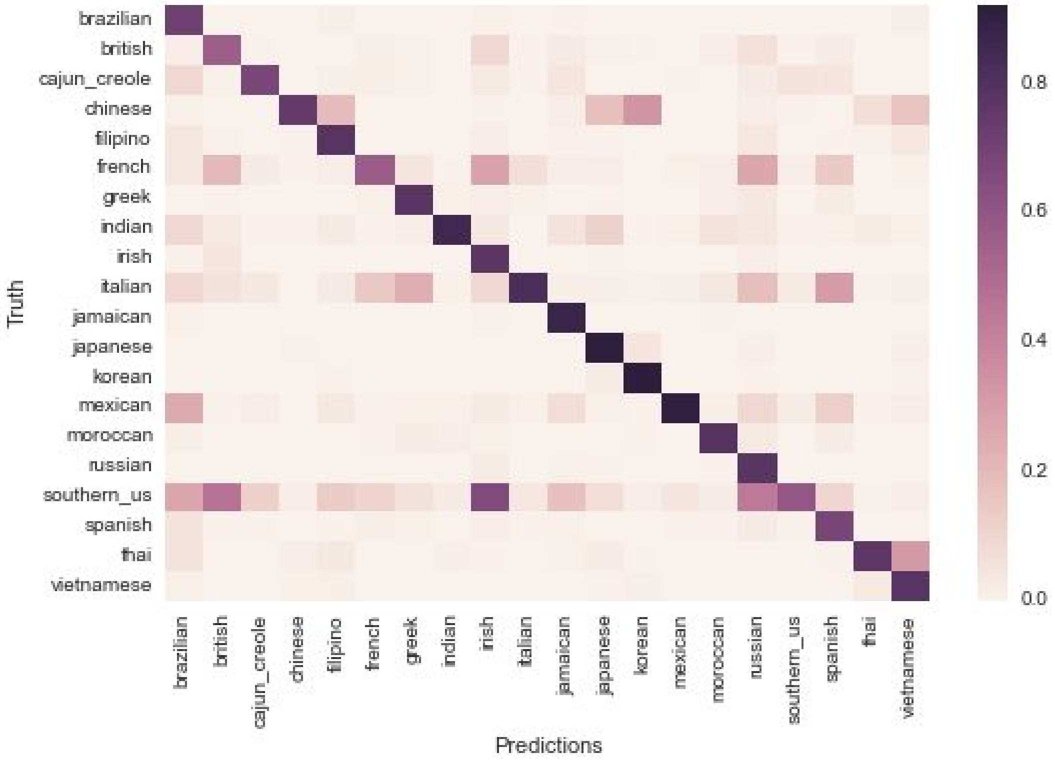
- Null Model: 19% (predict Italian for everything)

Heatmap Interpretation

- Heatmaps are adjusted to reflect different proportions of each cuisine
- Dark purple diagonal reflects observations where our predictions match the truth--where we are right
- Darker vertical stripes indicate cuisines which are overpredicted.
- Darker horizontal stripes indicate cuisines with which we have particular difficulty.
- Included percentages are test set accuracies.

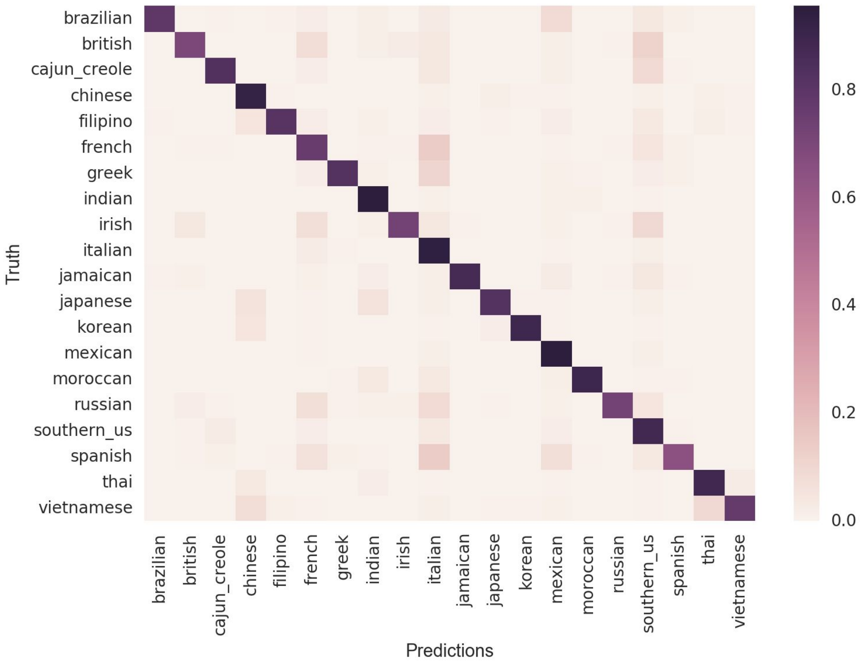
Naive Bayes

- 73% accuracy on test set

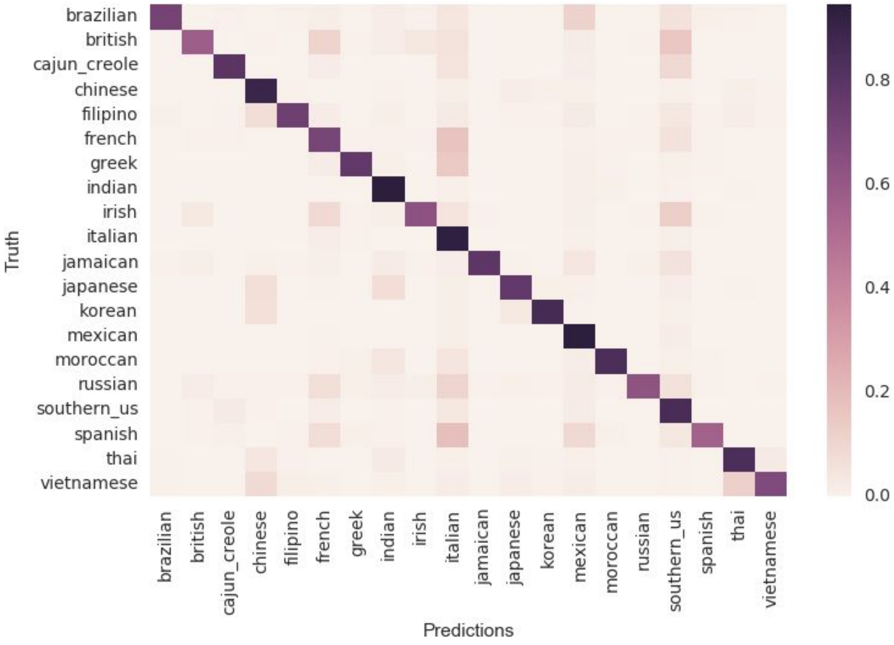


Linear Models

- Logistic Regression: 78%

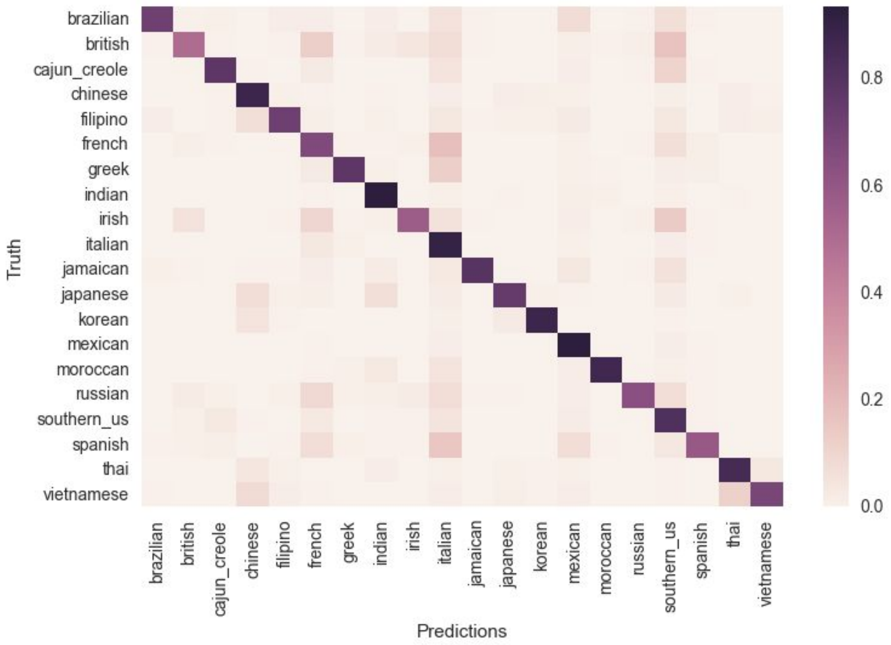


- Support Vector Machines: 78%

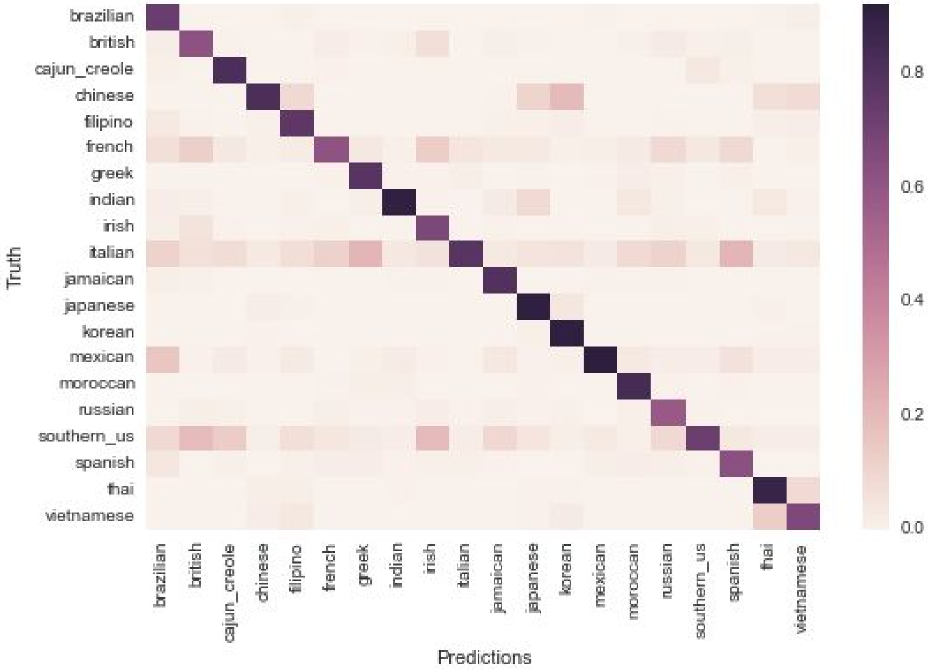


Dimensionality Reduction Methods

- Latent Semantic Indexing: 71%

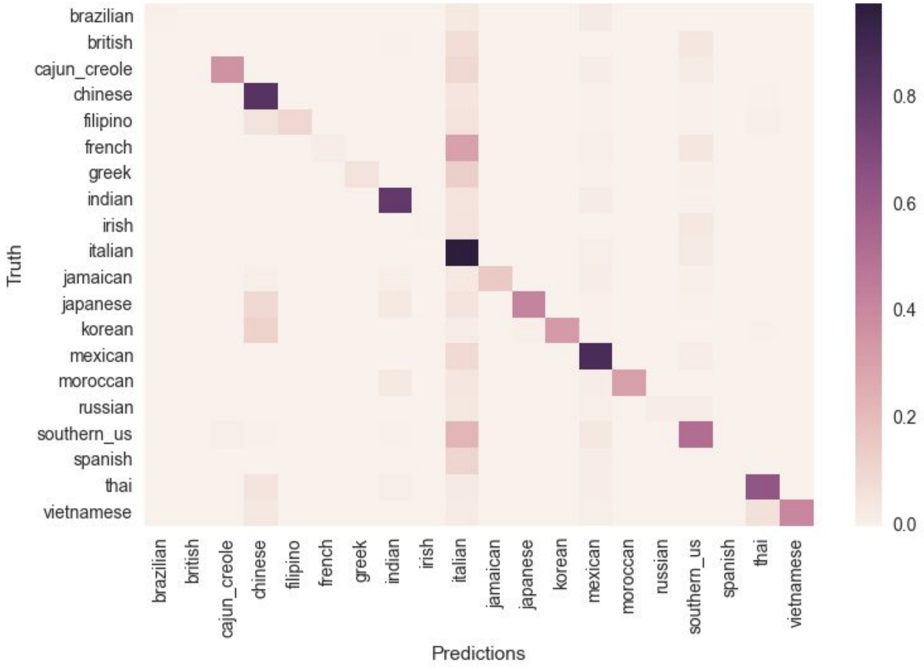


- Linear Discriminant Analysis: 74%

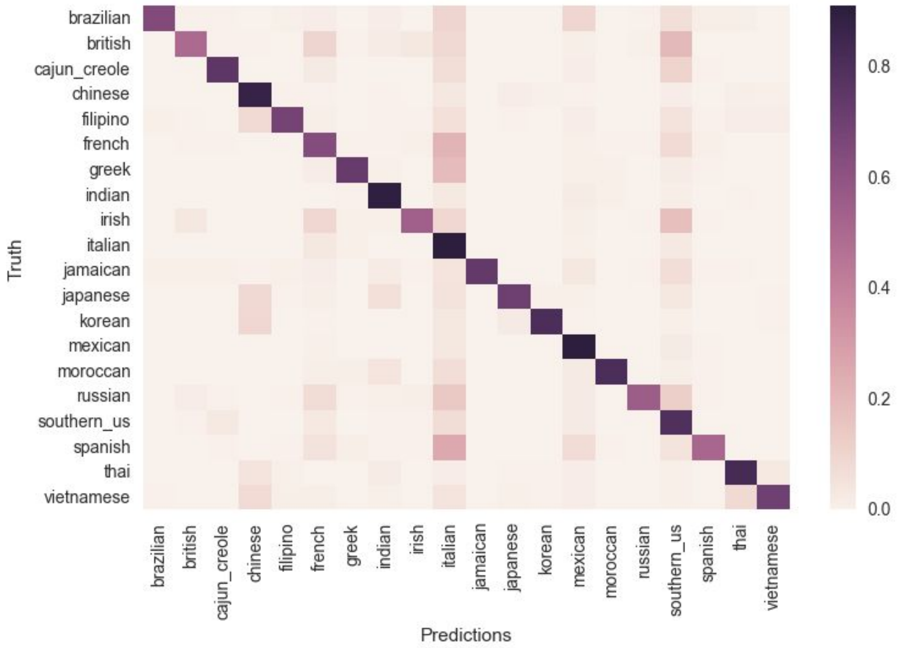


Tree Based Ensembles

- Random Forests: 55%

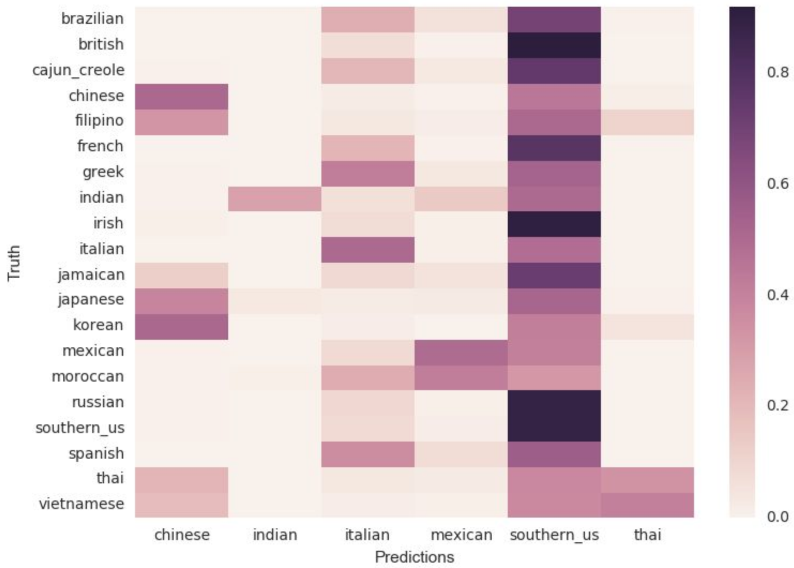


- Gradient Boosting: 76%



Cooperative Ensemble

- Use decision tree to split dataset into subsets (10 leaves)
- Fit and tune models for each subset
- 35% accuracy--each model always predicted majority class of its node



Competitive Ensemble

- Individual model accuracies vary between 71% and 78%
- 67% unanimity
- Algorithm: Pick mode. Pick higher prior if tie
- 79% accuracy

XGB	76%
LSI->Logistic Regression	71%
Logistic Regression	78%
SVM	78%
Naïve Bayes	73%
LDA	74%

Takeaways

Conclusions

- Yes, with roughly 78% accuracy
- Ingredient amounts and order would be useful
- Top competitors are about 82%
- Are cuisine distinctions meaningful when we’re recommending dishes?

Future Plans

- Custom weights on classes
- Neural networks
- Tune competitive ensemble

Shoutouts

- Scikit-Learn
- Scipy Sparse Matrices
- XGBoost
- Seaborn
- Tagul