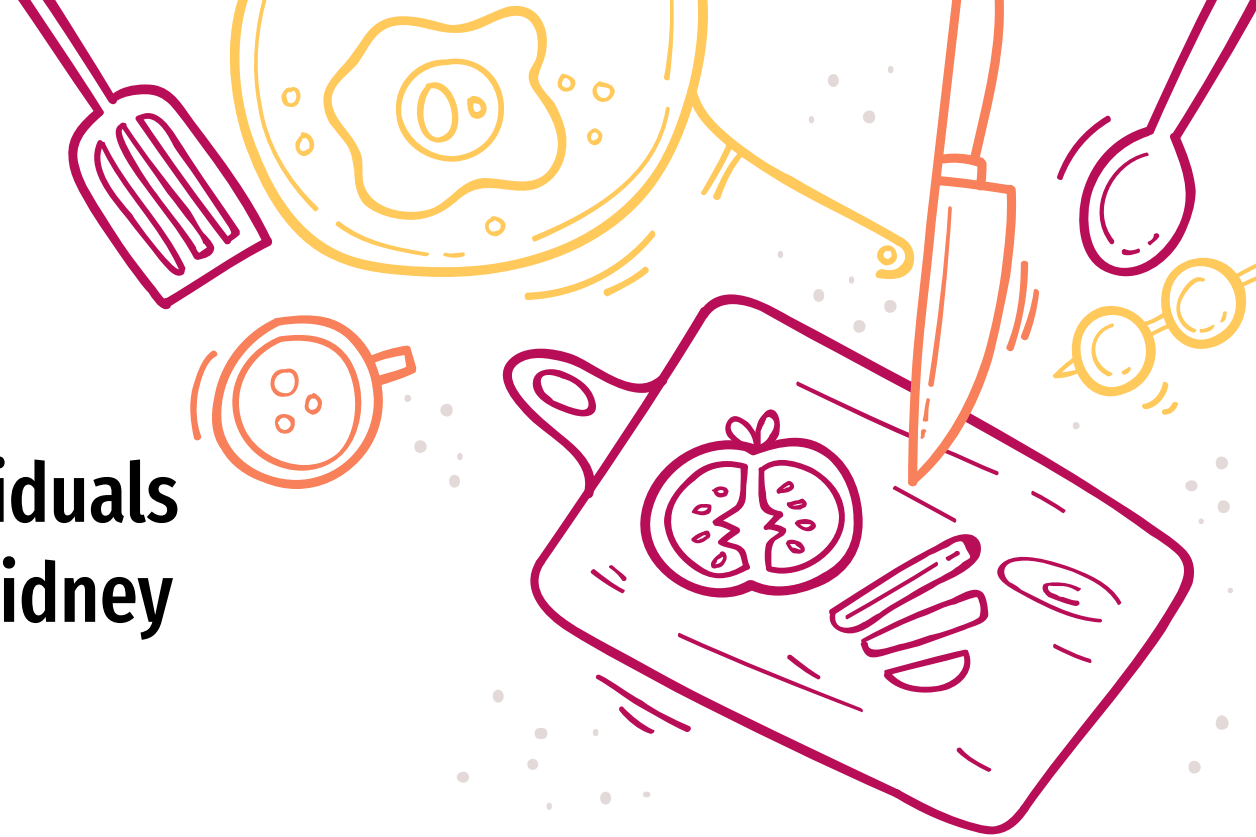
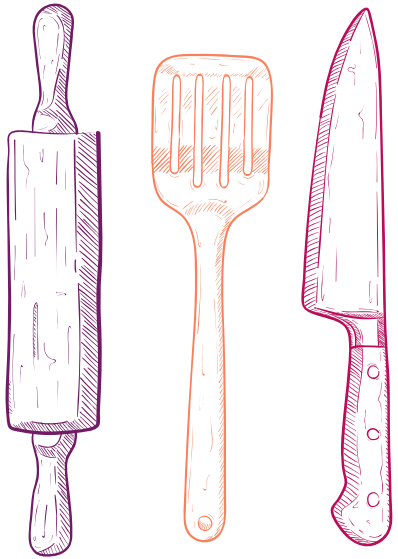
Predicting Recipe Suitability for Individuals Managing Chronic Kidney Disease (CKD)



Motivating Factors



CKD Diet

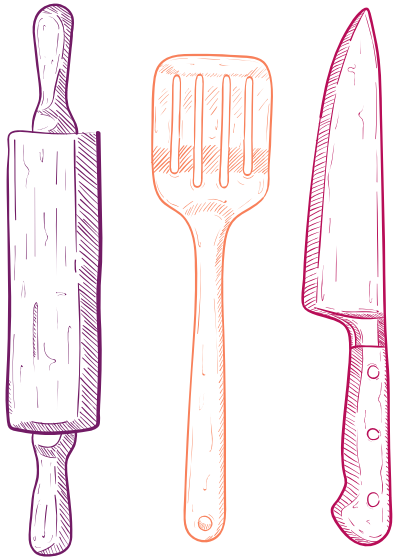
Managing CKD requires nutritional intake vigilance, and nutritional information is not guaranteed

Recipe Availability

Recipe collections exist, but are smaller in number when compared to recipe collections widely available

Professional Review

Curating recipes is a time consuming process dependent on a limited body of professionals



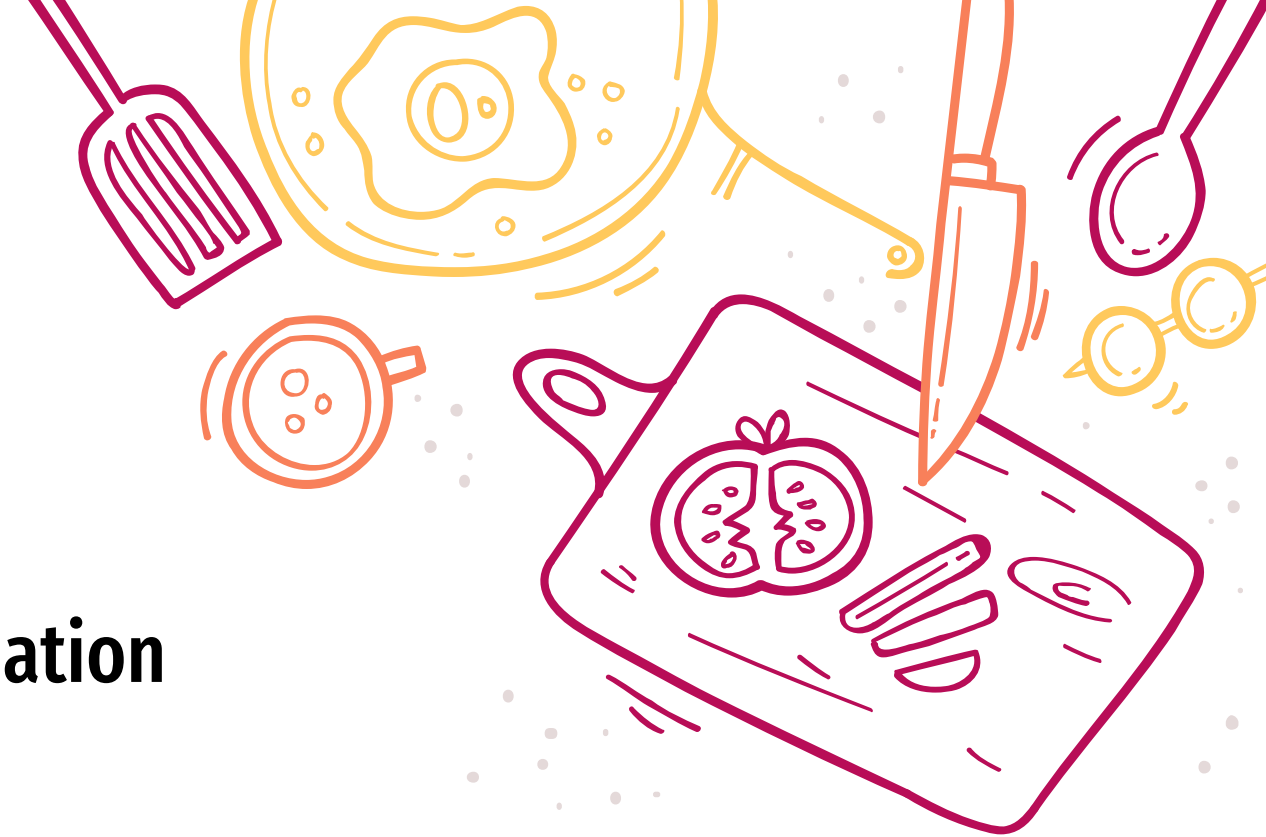
Problem Statement

Chronic kidney disease (CKD) affects >10% of the general population and has shown an increase in associated deaths in the last 20 years. Nutritional intake is an important aspect of managing progression of the disease, but identifying the appropriateness of recipe ingredients can be challenging for individuals and time consuming for professionals.

This project sought to answer the question of whether recipe ingredients could be distinguished by their similarity or dissimilarity to existing curated renal-friendly recipes using natural language processing (NLP) and classification models.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9073222/

Definitions and Background Information



What is Chronic Kidney Disease?

Also called: CKD, acute kidney disease, chronic renal failure, end stage renal disease, ESRD

Chronic kidney disease occurs when kidneys are no longer able to clean toxins and waste product from the blood and perform their functions to full capacity. This can happen all of a sudden or over time.

The severity of CKD is considered Early or Late stage, with 5 substages. Level of care and management changes throughout stages.

This project focuses on nutritional recommendations that are appropriate for all individuals with CKD.

Stages of Kidney Disease

Early Stage

Stage 1 - Normal function
Stage 2 - Mild CKD
Stage 3a - Moderate CKD
Stage 3b - Moderate CKD

Late Stage

Stage 4 - Severe CKD
Stage 5 - End Stage

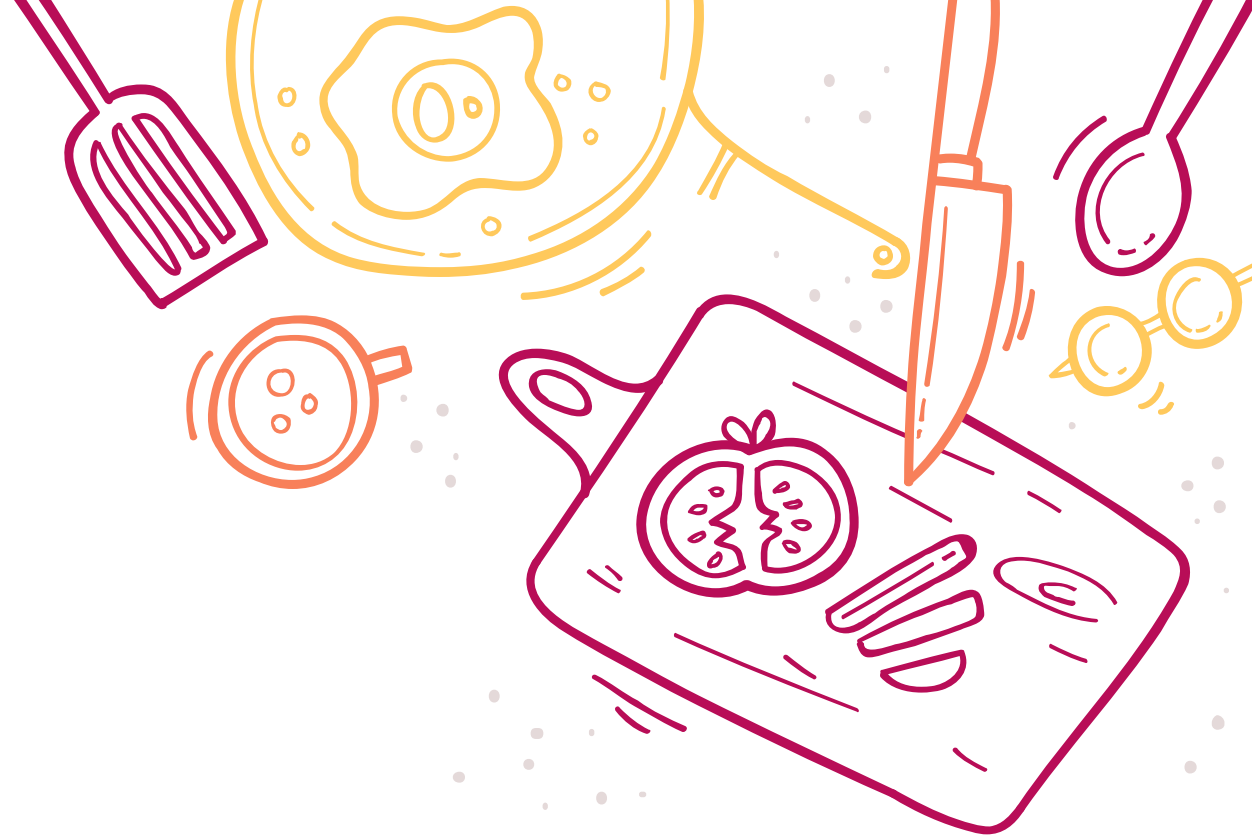
https://www.davita.com/education/kidney-vocabulary/chronic-kidney-disease
https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/eating-nutrition

Nutritional Guidelines

Sodium	Most important for managing CKD. Affects blood pressure. Lower is better.
Protein	Using protein produces waste that the kidneys must filter. Lower is better*
Saturated Fat	Important for heart health, a parallel objective for managing CKD. Lower is better.
Phosphorus and Potassium	Appropriate levels important for later stages. **Not considered** in this project.

*There are high-protein collections of curated renal-friendly diets. While this may be appropriate for some, most findings concluded that low amounts of protein were beneficial for managing kidney health

Data Acquisition



Recipe Extraction - Kidney Friendly

KF		
	Data Category	Curated kidney-friendly recipes
	Specific Data	Recipe ingredients, serving size, nutritional info
	Method	Web scraping (BeautifulSoup, Selenium)

Kidney Community Kitchen

- Curated by the Kidney Foundation of Canada
- 167 recipes (high protein omitted)

KidneyKitchen

- Curated by the American Kidney Fund
- 667 recipes (high protein omitted)

Recipe Extraction - “Other” Recipes

Other		
	Data Category	High sodium, high protein, or high fat recipes
	Specific Data	Recipe ingredients, serving size, nutritional info
	Method	Kaggle datasets

Food.com Recipes and Interactions

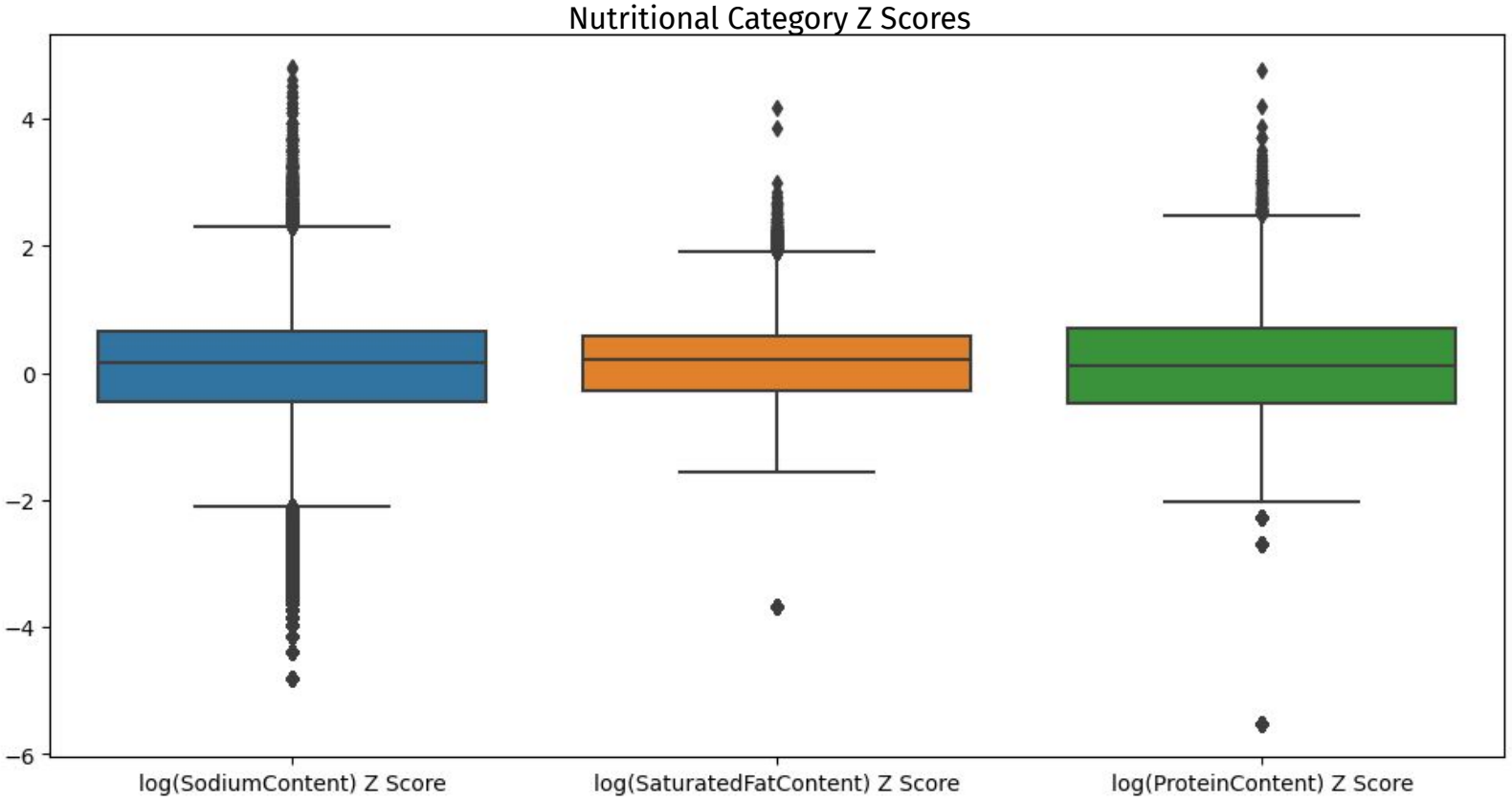
- Obtained as part of a Kaggle dataset
- Raw ingredients as string
- No nutritional information
- 500,000+ recipes

+

Food.com Recipes and Reviews

- Obtained as part of a Kaggle dataset
- Contained nutritional information
- 400,000+ recipes

“Other” Data EDA - Identifying Outlier



Filtering “Other” Recipes to Find Unfriendly Entries

Sodium

Derived from Kidney Institute’s designation of high sodium being > 20% of the FDA recommended intake per serving

Protein

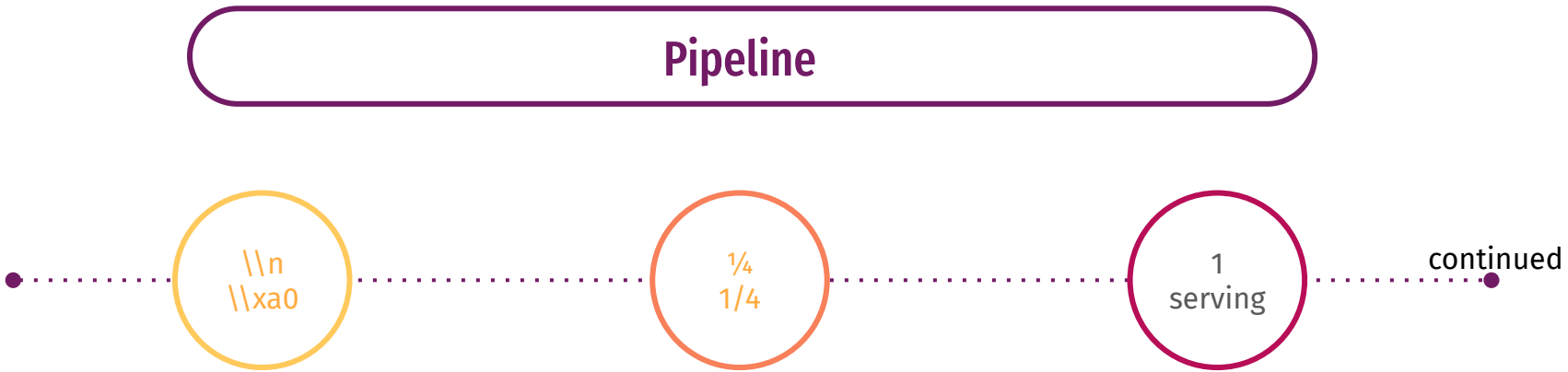
Derived from a calculation involving the median weight of a U.S. male, recommended protein intake for CKD sufferers, and number of meals per day

Saturated Fat

Derived from recommended sat. fat intake for general population, caloric content per g/fat, numbers of meals per day

Data Preparation

Data Cleaning / Preprocessing



Strip characters

Eliminate special and whitespace characters, quotes, brackets, etc

Convert fractions

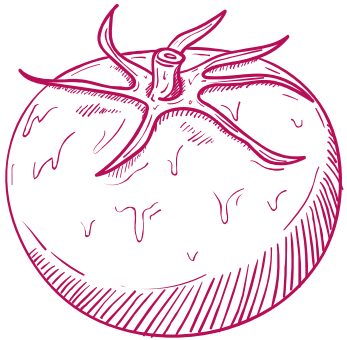
Replace vulgar fractions with decimals. Targeted standalone unicode characters and separated characters

Divide by serving size

Divide all quantities so that all recipe ingredients were in terms of one serving

Final step - standardization

In order to aid the model in understanding the significance of quantities, ingredient measurements were converted to grams (as best as possible)



“2 pounds of tomatoes”

907 grams tomatoes

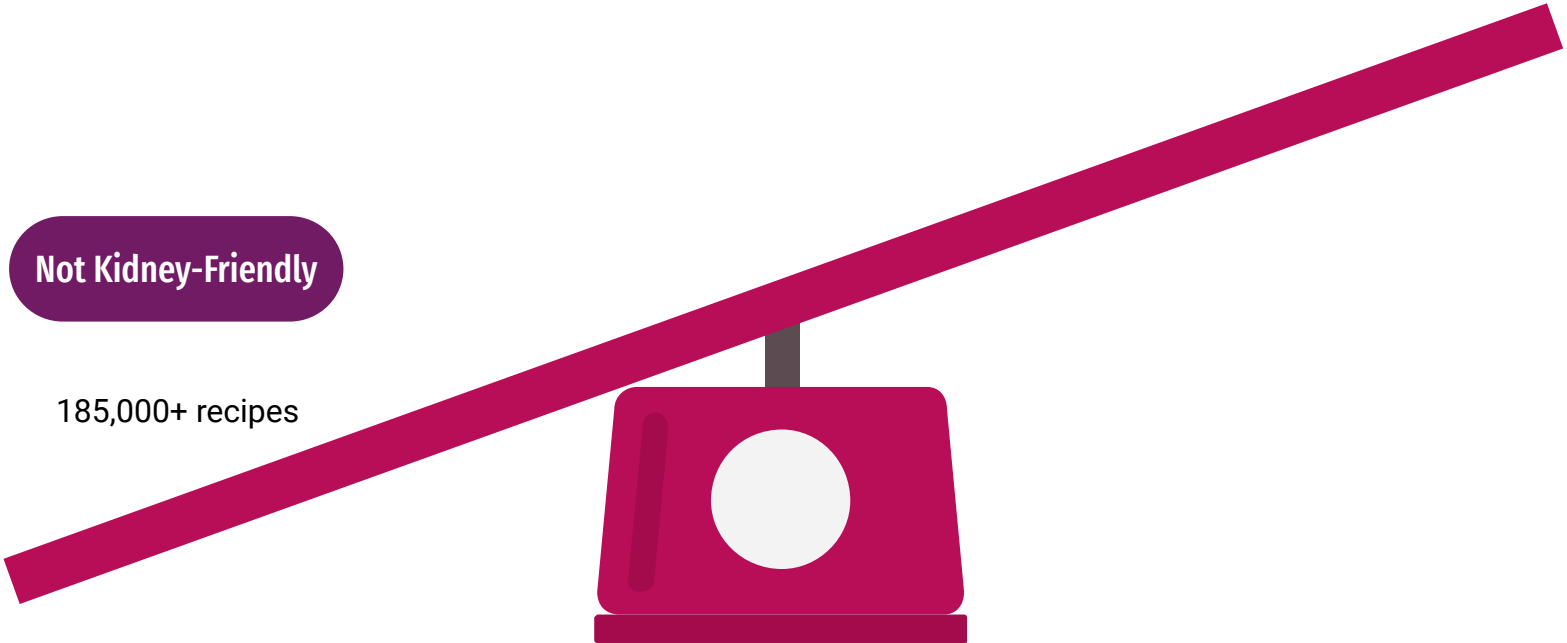
227 grams mushrooms

“Mushrooms: 8 oz”

Class Imbalance

Kidney-Friendly

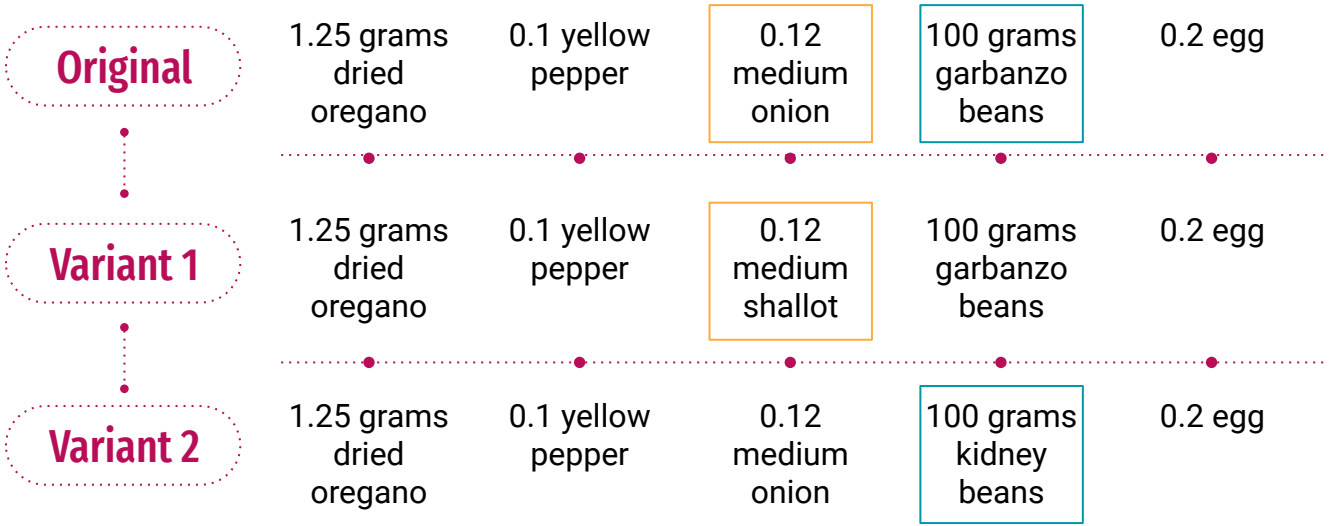
~800 recipes



Not Kidney-Friendly

185,000+ recipes

Data Augmentation Via Recipe Variants



Class Imbalance

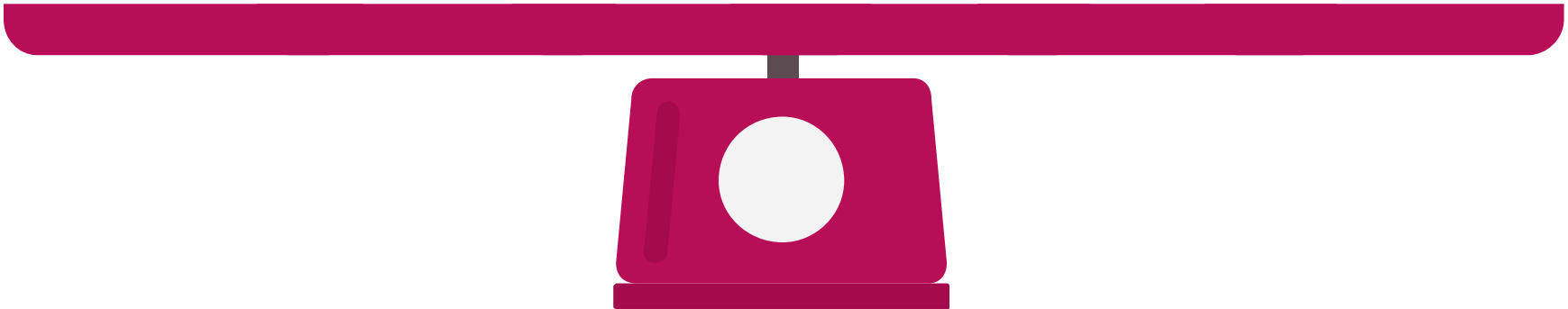
(Augmented) Kidney-Friendly

~8,600 recipes

(Baseline accuracy: 53.5%)

(Sample of) Not Kidney-Friendly

10,000 Recipes



Modeling

Model Details and Performance

Model 1

Type:

Recurrent Neural Network
using trainable Embedding, Bidirectional, Dropout, and Dense layers

Input

Standardized ingredients and measurements run through a TextVectorization Object

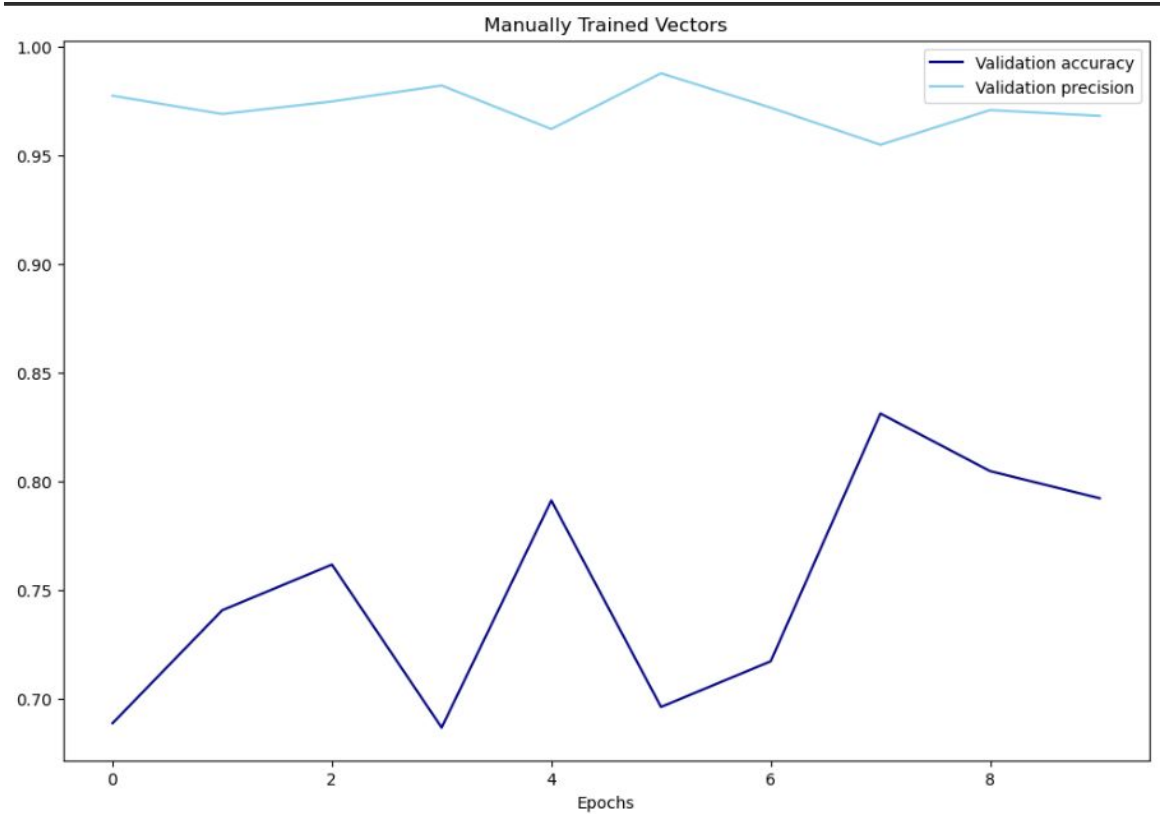
Embedding

5,120,000 trainable params

Precision / Accuracy

97% / 80%

Model Details and Performance



Model Details and Performance

Model 2

Type:

Recurrent Neural Network
using pretrained Embedding, regularized Bidirectional, Dropout, and Dense layers

Input

Standardized ingredients and measurements run through a TextVectorization Object

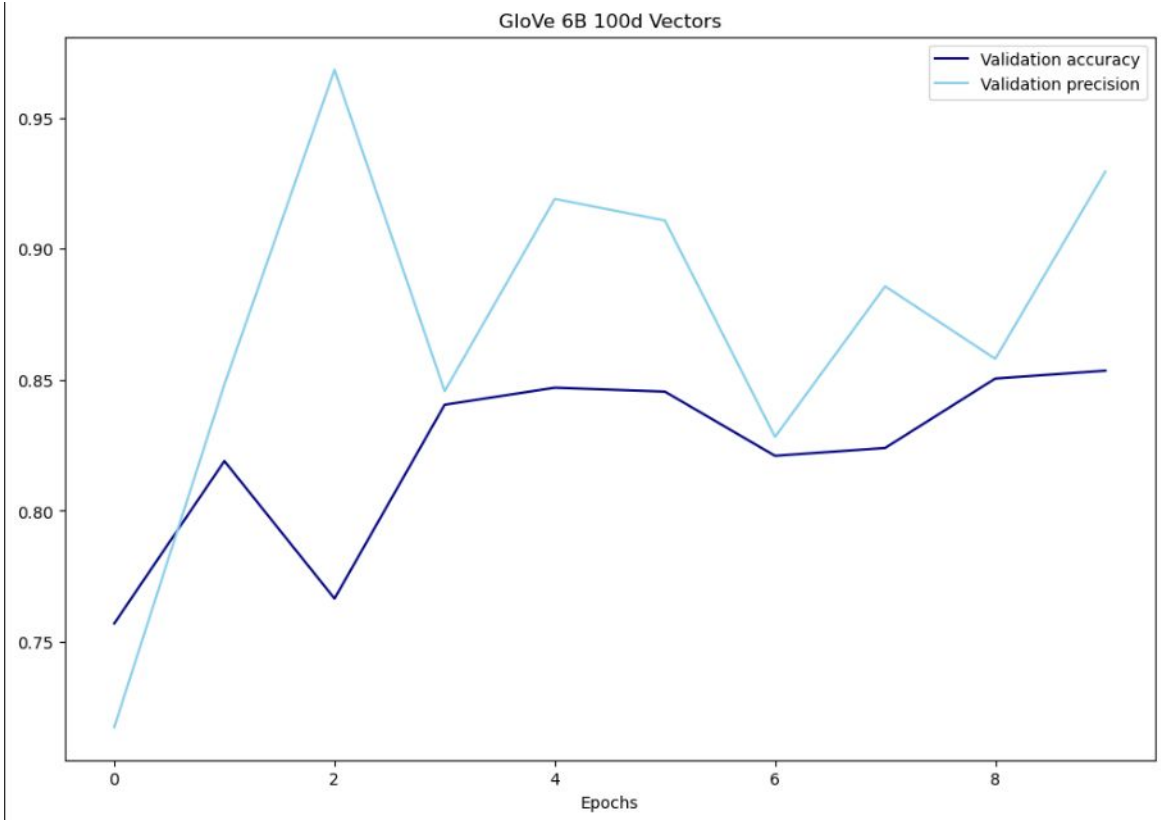
Embedding

Global Vectors (GloVe) 6B Tokens 100 dimensional

Precision / Accuracy

93% / 85%

Model Details and Performance



Model Details and Performance

Model 3

Type:

Recurrent Neural Network
using pretrained Embedding, regularized Bidirectional, Dropout, and Dense layers

Input

Standardized ingredients and measurements run through a TextVectorization Object

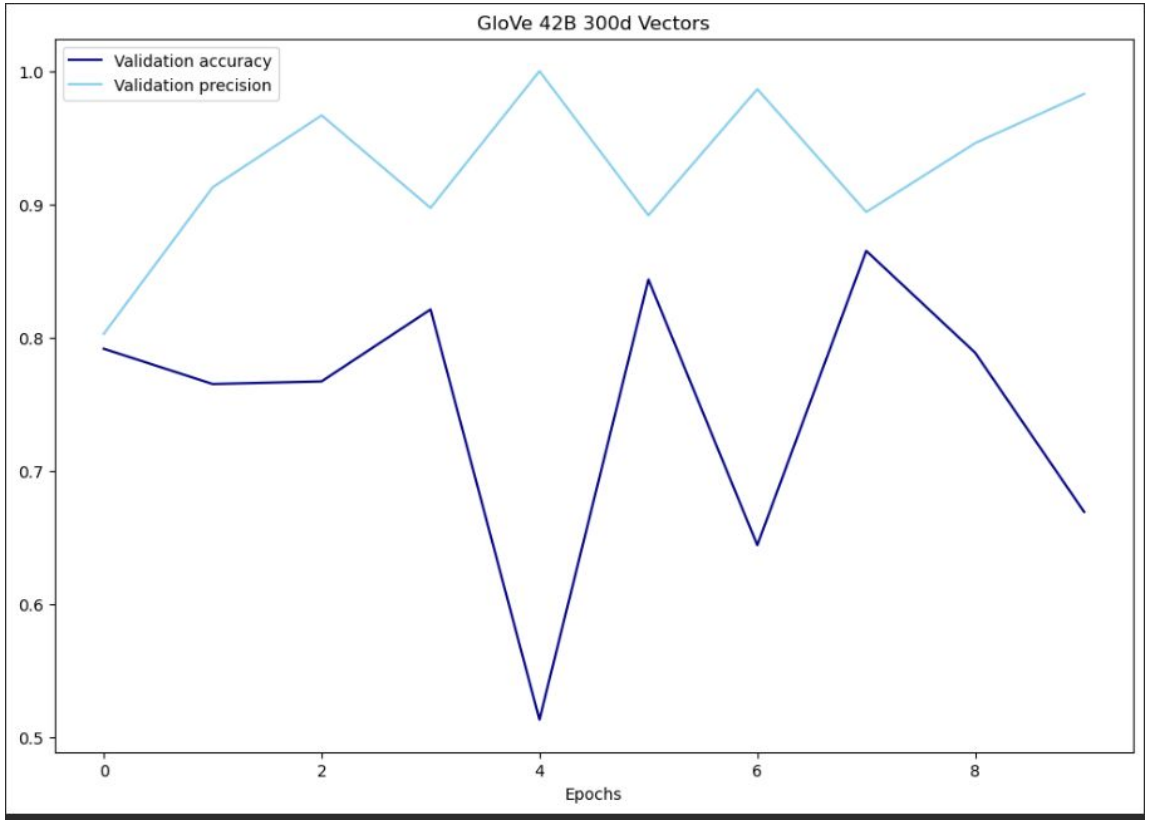
Vocabulary

Global Vectors (GloVe) 42B Tokens 300 dimensional

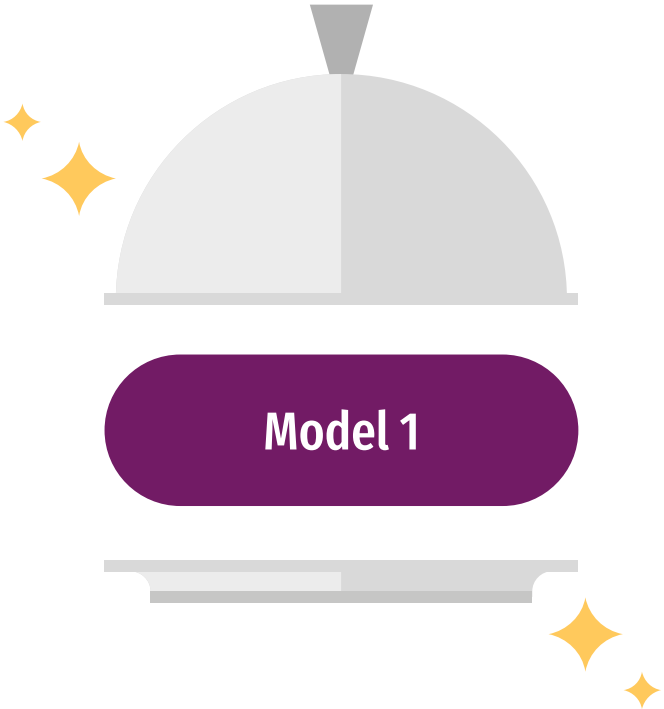
Precision / Accuracy

97% / 77%

Model Details and Performance



Result



Precision

- 97%

Accuracy

- 80%

Reasoning

- Chosen for its high precision and respectable accuracy

Limitations

With additional time, improvements could originate in these areas:

Data

Increasing the amount of genuine recipes, or even including preparation and descriptions

Expertise

Information from subject matter experts could craft better thresholds, recipes, and shed light on things like Bayesian Error

Standardization

Standardizing ingredient measurements could be more elaborate, perhaps even through a other models

Technique Variety

Exploration of other model types and NLP techniques could potentially yield more favorable results

Citations and Acknowledgements

Special Thanks:

- This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**
- Tim Book -for all data science instruction, but in the context of this project: data augmentation and RNNs with NLP
- Rowan Schaefer for assistance and willingness to be a sounding board
- American Kidney Fund and Kidney Foundation of Canada for their curation of kidney-friendly recipes
- Jeffrey Pennington, Richard Socher, Christopher D. Manning for use of Global Vectors for Word Representation (GloVe)

Data:

- **https://www.kaggle.com/datasets/irkaal/foodcom-recipes-and-reviews**
- **https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions**
- **https://kitchen.kidneyfund.org/find-recipes/**
- **https://www.kidneycommunitykitchen.ca/kkcookbook/recipes/**

External Tools & Information:

- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. [pdf] [bib]
- Brownlee, Jason. (2020, August 25). Use Early Stopping to Halt the Training of Neural Networks At the Right Time. MachineLearningMastery.com. https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/\

Dietary Information

- National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Sodium Tips for People with CKD. Accessed December 5, 2023. https://www.niddk.nih.gov/-/media/Files/Health-Information/Health-Professionals/Kidney-Disease/SodiumTipsforPeopleCKD EN.pdf
- National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Eating Right for Chronic Kidney Disease. Accessed November 29, 2023. https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/eating-nutrition
- Ko, Gang Jee, et al. "Dietary Protein Intake and Chronic Kidney Disease." Current Opinion in Clinical Nutrition and Metabolic Care 20.1 (2017): 77-85. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6855949/
- Mayo Clinic. How to track saturated fat. [Online]. Rochester, MN: Mayo Clinic; 2023 [Updated March 03, 2023]. Available from: https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/saturated-fat/art-20045858