

# Discovering Healthier Recipes

using natural language processing and probabilistic big data structures



**RIT**

*Matthew Ormson*  
MTO4125@RIT.EDU

*Dr. Christopher Homan*  
CMHVCS@RIT.EDU

## Introduction

## The Problem

With finding healthy recipe alternatives:

1. Finding recipes takes time and manual research
2. Similarity between recipes is solely based on name/description
3. Comparing nutrition requires manual computation and is incomplete

## The Solution

Design a system that can:

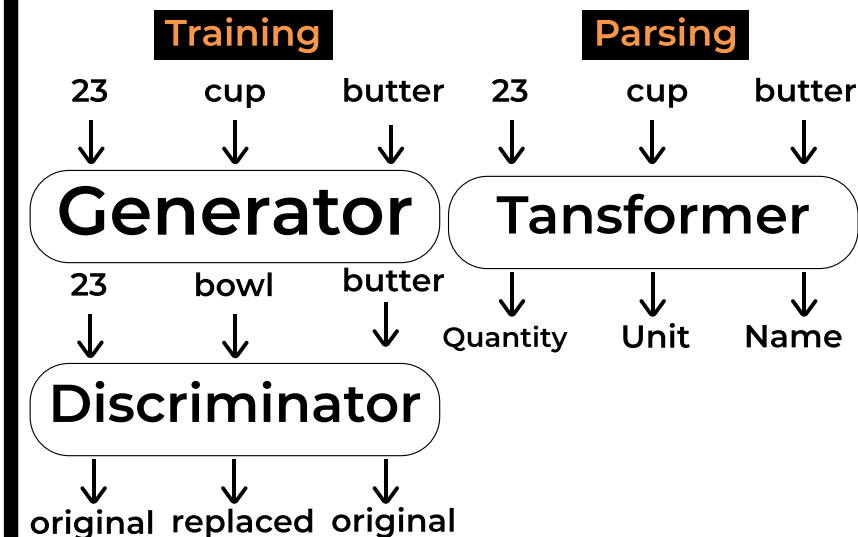
1. Efficiently store a large number of recipes
2. Quickly find similar recipes based on ingredients
3. Easily compare any nutritional measurement between similar recipes

## Methodology

- 1,029,720 recipes scraped from the web
- 52 Nutritional Measurements from USDA
- 14,336 Ingredient Substitutions from The Cook's Thesaurus

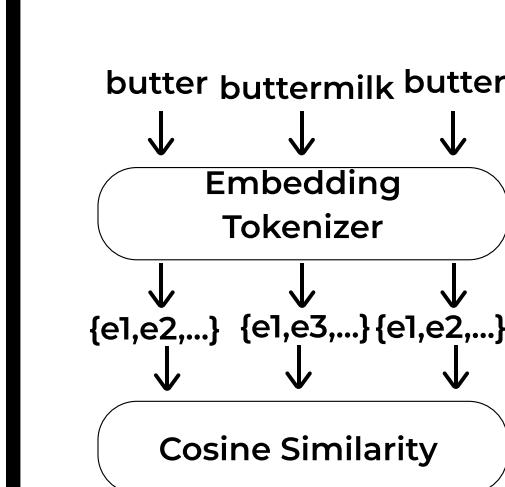
### Data Parsing

Electra



### Data Matching

Cosine Similarity



### Recipe Similarity

Jaccard Similarity

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

MinHash

Jaccard + O(nk)

Locality Sensitive Hashing

Similar items are more likely to collide into the same hashing bucket

MinHash LSH Forest

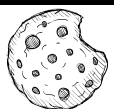
MinHash + LSH + smart indexing = top K similar

### Comparing Nutrients

Sql Queries

```
SELECT ingredients.RecipeID, ingredients.Name,
SUM(ProteinCalculated) AS ProteinCalculated
FROM (
SELECT r.RecipeID, r.Name,
i.Quantity * n.Protein as ProteinCalculated
FROM Data.Ingredients i
INNER JOIN Data.Nutrition n
ON n.NutritionID = i.NutritionID
INNER JOIN Data.RecipeSimilarities sr
ON sr.SimilarRecipeID = i.RecipeID
AND sr.RecipeID = 'N')
INNER JOIN Data.Recipes r
ON r.RecipeID = sr.SimilarRecipeID
) as ingredients
GROUP BY ingredients.RecipeID, ingredients.Name
```

## Example



### Blueberry Coconut Cookies

- 23 cup butter - 2 cups flour
- 24 cup sugar - 1 cup coconut
- 12 cups milk - 1 cup blueberries

Name	Unit	Quantity
butter	23	cup
sugar	24	cup
milk	12	cups
flour	2	cups
coconut	1	cup
blueberries	1	cup

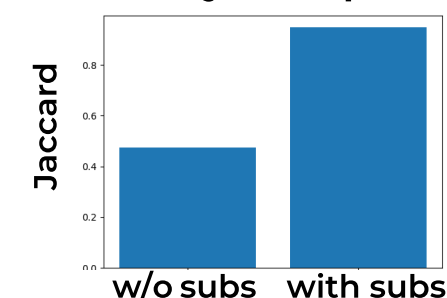
Name <sub>1</sub>	Name <sub>2</sub>	Cosine Similarity
butter	butter	1
butter	buttermilk	0.43
butter	school bus	0

Recipe	Jaccard
Blueberry Coconut Cookies	1
Low Sugar Blueberry Coconut Cookies	0.9453125
Honey-Glazed Buttermilk Oatmeal Coffee Cake	0.5078125
Cinnamon Banana Bread	0.484375
Diabetic Blueberry Muffins	0.4609375
Blueberry Muffins (With Splenda)	0.4609375
Banana Bread	0.3984375
Sherry Pumpkin Bread	0.3359375
Apple Walnut Upside-Down Cake	0.328125
Sticky Toffee Pudding	0.3125

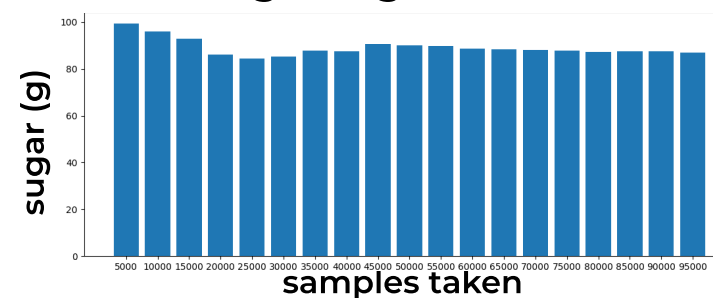
Recipe	Sugar(g)
Cinnamon Banana Bread	174.94
Banana Bread	186.34
Sherry Pumpkin Bread	204.17
Low Sugar Blueberry Coconut Cookies	251.25
Blueberry Muffins (With Splenda)	254.72
Diabetic Blueberry Muffins	254.72
Blueberry Coconut Cookies	302.61
Sticky Toffee Pudding	342.43
Apple Walnut Upside-Down Cake	457.64
Honey-Glazed Buttermilk Oatmeal Coffee Cake	494.29

## Results

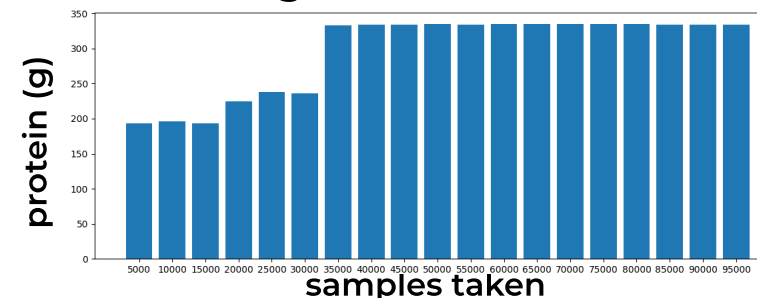
### Similarity Comparison



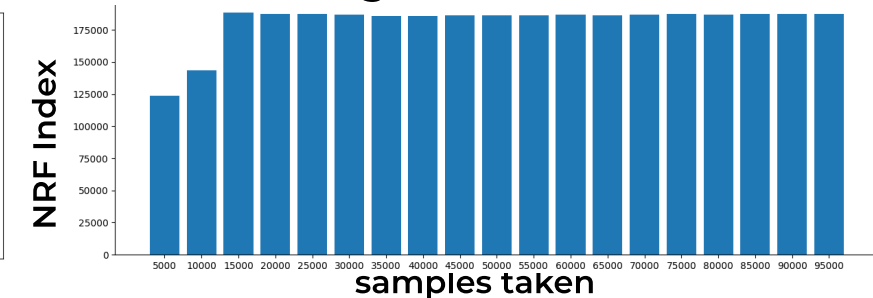
### Average Sugar Decrease



### Average Protein Increase



### Average NRF Increase



### NRF Index

