# Where to Build Food Banks and Pantries: A Two-Level Machine Learning Approach

Gavin Ruan, West Lafayette Jr./Sr. High School

## 1. Introduction

**Motivation and Background:**

- In the United States, over **44 million** people suffer from food insecurity, **13 million** of whom are children.
- Food banks are critical for these people's nutrition and health (**49 million** turned to food programs in 2022).
- Last year, I created an application using the K-Means clustering algorithm. While this improved food bank locations, it didn't factor in important considerations such as roads or resources such as food pantries.

**Project Goal:**

- Improve the initial machine learning approach for finding optimal locations by **considering roads** and creating a **two-level system that has food banks and pantries.**

## 2. Methodology

**Optimization:**

- Find the **food bank and pantry** locations with the shortest total **road distance** to all served households

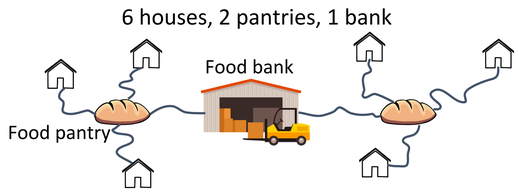
**Datasets:**

- Datasets consisted of Indiana houses from 2020 GIS data

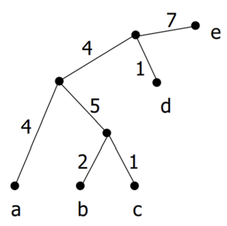
OSRM
OPEN SOURCE ROUTING MACHINE

**OSRM:**

- Open source geospatial data
- Requires data to be scaled down in size




| M | a | b | c | d | e |
|---|---|---|---|---|---|
| a | 0 | 11 | 10 | 9 | 15 |
| b | 11 | 0 | 3 | 12 | 18 |
| c | 10 | 3 | 0 | 11 | 17 |
| d | 9 | 12 | 11 | 0 | 8 |
| e | 15 | 18 | 17 | 8 | 0 |




**Distance/Dissimilarity Matrix**

**K-Medoids Method:**

- Unsupervised clustering algorithm
- Heuristic approach to solve facility location problems quickly using a distance matrix

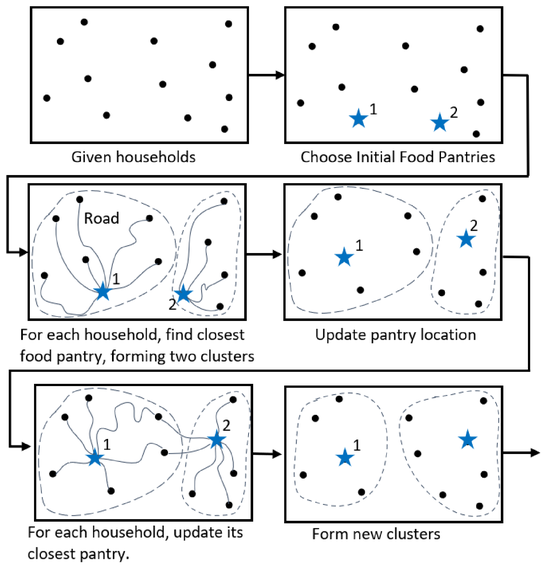
**Comparisons:**

- Generated 11 food banks and 176 total pantries to compare to an equal number of real ones

**Performance:**

- Jupyter notebook's time function for computational cost
- OSRM road distance converted to miles for distance measurements

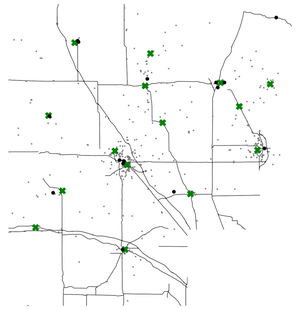
Example flow chart for 12 houses and 2 food pantries



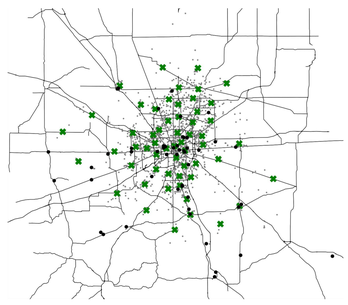

## 3. Results

### Pantry-Household Level

● Real Food Pantries ✖ AI Generated Pantries



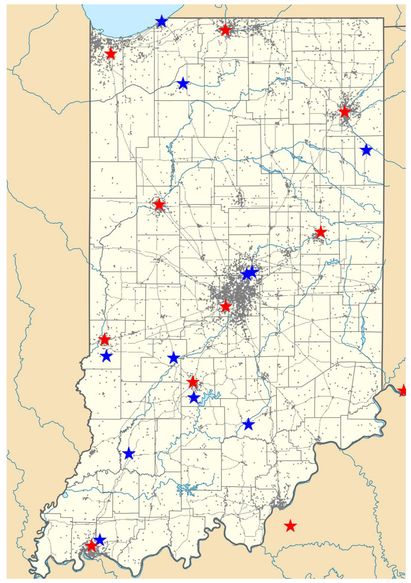
Lafayette
Houses: 456
Number of pantries: 15
Original distance: 9.34 mi.
AI distance: 6.37 mi.
Distance saving: 2.97 mi.



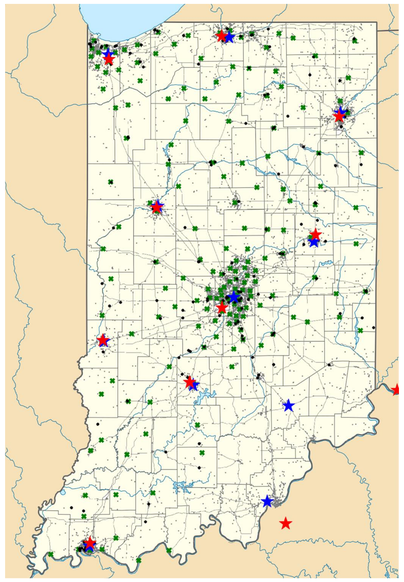
Indianapolis
Houses: 1718
Number of pantries: 57
Original distance: 4.41 mi.
AI distance: 2.25 mi.
Distance saving: 2.16 mi.

### Food Bank-Pantry Level

★ Real Food Banks ★ AI-Generated Food Banks



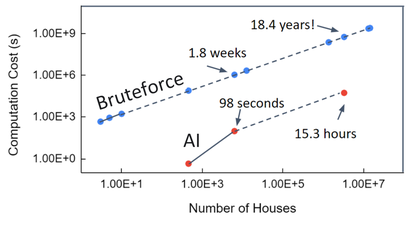
Initialization

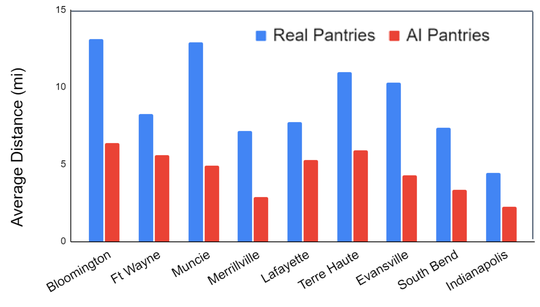


Converged Results

**Large Distance Savings between Pantries and Households:**

- 6293 Houses
- Clusters ranged from 270 to 1700 houses
- **3.52 average miles saved per household**
- **22,181.423 total miles saved**
- **Average saved miles ranged from 2.17 to 8 miles in different cities**

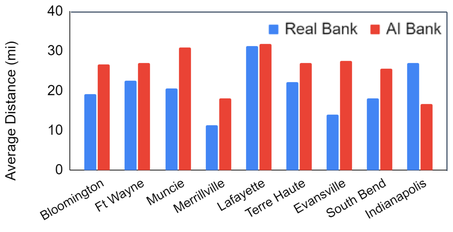



| # Houses | Bruteforce | AI |
|---|---|---|
| 6293 houses | 1.8 weeks | 98 seconds |
| ~3.3 million houses | 18.4 years | 15.3 hours |

**Small Distance Penalty between Food Banks and Pantries:**

- 176 pantries
- 1.56 average miles penalty per pantry
- 273.75 total miles penalty




## 4. Conclusions

- Results show that my two-level machine learning approach is able to consider real roads and generate a set of food banks and pantries both extremely quickly, and with more optimized locations than current existing ones.
- Current layouts prioritize food pantry proximities with food banks
- Contrary to the status quo, AI has showed that the planning strategy needs to be changed to prioritize households

## 5. Future Work

- Consider capacity of food banks
- Try a bottom-up approach with the food bank placements
- Expand dataset size and include different weights for houses (income range, socio-economic data, etc.)

## 6. References

Celik Turkoglu, D., Erol Genevois, M. A comparative survey of service facility location problems. Ann Oper Res 292, 399–468 (2020). https://doi.org/10.1007/s10479-019-03385-x

Héctor J. Carlo, Francisco Aldarondo, Priscilla M. Saavedra & Silmarie N. Torres (2012) Capacitated Continuous Facility Location Problem With Unknown number of Facilities, Engineering Management Journal, 24:3, 24-31, DOI: 10.1080/10429247.2012.11431944

Shih, H. (2015) Facility Location Decisions Based on Driving Distances on Spherical Surface. American Journal of Operations Research, 5, 450-492. doi: 10.4236/ajor.2015.55037.

Varghese, S.; Gladston Raj, S. Clustering Based Model For Facility Location In Logistic Network Using K-Means. Int. J. Sci. Invent. Innov. 2016, 1, 26–32.