

Machine Learning Based Identification of Food Desert Defect in Urban Development

Yiyuan Dorothy Zhao
Thesis Defense
Advisor: Dr. Sunnie Chung

Cleveland State University
6 May 2020

Agenda

- | | |
|--|-----------------------|
| I. Introduction | VI. Classification |
| II. Related Work | VII. Experiments |
| III. Data Description &
Preprocessing | VIII. Web Application |
| IV. Framework | IX. Conclusion |
| V. Methodology | X. Future Work |

1

Introduction

Background, motivation, problem statement, objectives



A **Food Desert** is a geographic location of blocks of neighbors where there is a dearth of available fresh, nutritious food.

Problem of Food Desert:

Food deserts – a highly relevant problem that indicates an interface between poverty, lack of food access, and poor health conditions



Background

- 20 million Americans live in food deserts
- Factors include transportation, socioeconomic status, price & presence of fresh food, location of grocery stores
- Healthy Cleveland Initiative – HEAL map



Motivation

- Build a “Food Desert Identifier”
- Need improvement on existing methods to identify Food Deserts
- Superior methods to use: ML algorithms, Big Data Analytics
- Users: Government agencies for better urban planning, healthcare professionals, nonprofit organizations

Problem Statement

The complex concept of identifying food deserts has been a part of urban and health studies, but it can now be an opportunity for computer science to solve.

- Identifying related key factors
- Extensive big data preprocessing



Objectives

- Employing big data preprocessing techniques to integrate geospatial data and geographical neighborhood demographics from heterogeneous and complex big data
- Completing extensive research for a comprehensive feature selection process that identifies optimal feature sets for the food desert identifier
- Applying big data analytic methods to discover geospatial knowledge for relevant key features for the food desert identifier
- Employing Machine Learning classifiers to build accurate models to identifying food deserts

2

Related Work

Traditional approach, Geospatial
Information, Machine Learning

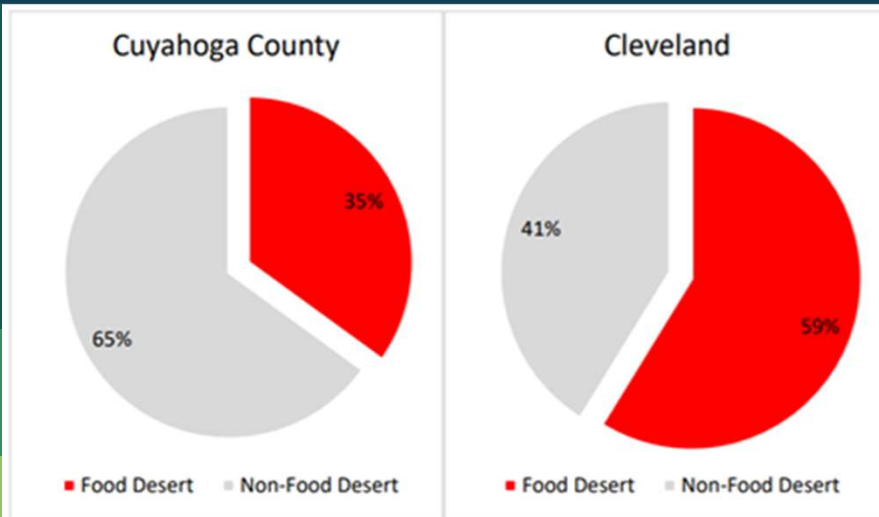


Traditional Approach

- 2017 Study: Evaluated domains of nutritious food access; 17 urban neighborhoods of Cleveland surveyed
- USDA ERS Food Access Research Atlas: spatial overview of food access indicators for low-income and other census tracts using different measures of supermarket accessibility
- Community Food Assessment: Cuyahoga County Planning Commission & Board of Health, 2018 Supermarket Assessment



Populations in Food Deserts



As of 2018, nearly 60% of Cleveland residents and approximately 35% of Cuyahoga County residents live in a food desert.



Studies with Geospatial Information

Geographic Information Systems (GIS)

- Studies in Connecticut, Alabama
- GIS mapping, in-person observations & surveys, ArcGIS software
- Spots of statistical significance with greater need of accessibility



Machine Learning Approach

U.C. Berkeley Food Desert Predictor by Graduate team

Predictive model anticipated food insecurity before it occurred, built upon USDA Food Access Research Atlas tool

- Completed in 2017, model prediction was at 80% accuracy



Comparison to Related Work

This thesis extends with:

- Additional key features: information from integrating big data & geospatial knowledge derived by data analytic methods
 - Prediction Accuracy Index (PAI)
 - Hotspot Analysis
 - Machine Learning approach

3

Data Description and Preprocessing

Collection and Description of Data,
Preprocessing and Integration

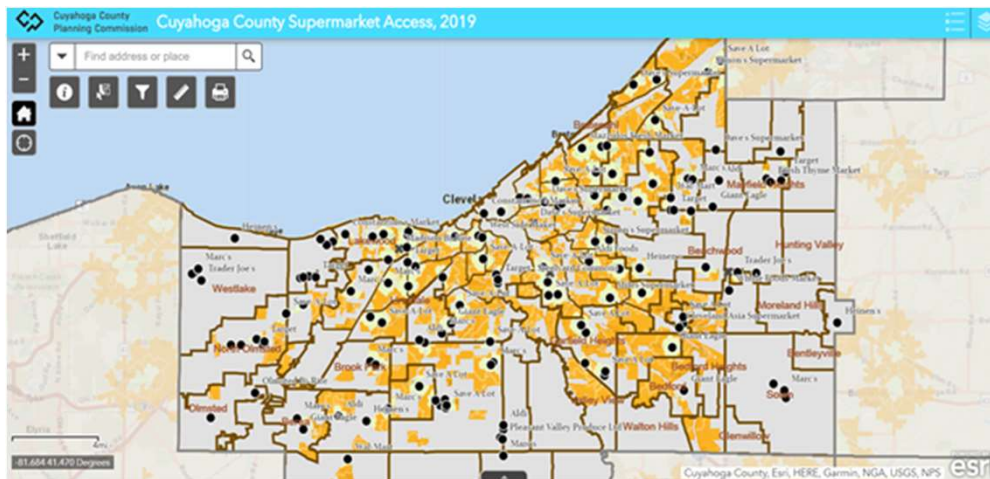


Collection of Big Data

- 3 Big Data Sources:
 - Land data – from CAMA System, Cuyahoga County
 - Property Info, Sales, Characteristic of Each Property
 - People data – from US Census Bureau, known as LODES
 - Residence Area Characteristics (RAC),
 - Workplace Area Characteristics (WAC)
 - Additional Geospatial Data from Multiple Sources
 - Food Desert Census Block dataset



Food Desert Census Block Data Based on Supermarket Access



Geospatial data of food deserts in census blocks collected by County Planning for Supermarket Assessment 2018 report

- Goal: displays supermarkets



Description of Data

Additional Geospatial Data Examples

- Poverty Threshold per Block Groups: Income is below 200% of federal poverty
 - Ex. Household income of \$48,000 for family of 4 crosses poverty threshold
- DriveMileZone: Distance from the center of each county block to a supermarket store location
 - Block: 0.03 square miles



Land Data

parcel	PARCEL_YEAR	linkid	cblgd	ciupdatedt	concl	cbasetyp	baseflr
107015	2013	8097392	1	6/30/2010	D	FUL	CNC
107017	2013	8097393	1	6/30/2010	D	FUL	CNC
107021	2013	8097397	1	1/18/2006	D	FUL	CNC
107022	2013	8097398	1		D	FUL	CNC
107023	2013	8097399	1	1/18/2006	D	SLB	CNC
107024	2013	8097400	1	1/18/2006	D	FUL	CNC
107027	2013	8097403	1	1/18/2006	D	FUL	CNC
107061	2013	8097436	1	1/18/2006	D	FUL	CNC
107077	2013	8097452	1	1/18/2006	D	PRT	CNC

Sample Commercial Characteristic Data

Data is divided by taxable land use; 7 characteristic files

Land Use data -
properties divided
into parcels of land
with ownership data



WAC Data Layout

Col	Variable	Type	Meaning
1	w_geocode	Char15	Workplace Census Block Code
2	C000	Num	Total number of jobs
3	CA01	Num	Number of jobs for workers age 29 or younger
4	CA02	Num	Number of jobs for workers age 30 to 54
5	CA03	Num	Number of jobs for workers age 55 or older
6	CE01	Num	Number of jobs with earnings \$1250/month or less
7	CE02	Num	Number of jobs with earnings \$1251/month to \$3333/month
8	CE03	Num	Number of jobs with earnings greater than \$3333/month
9	CNS01	Num	Number of jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)
10	CNS02	Num	Number of jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)
11	CNS03	Num	Number of jobs in NAICS sector 22 (Utilities)
12	CNS04	Num	Number of jobs in NAICS sector 23 (Construction)
13	CNS05	Num	Number of jobs in NAICS sector 31-33 (Manufacturing)
14	CNS06	Num	Number of jobs in NAICS sector 42 (Wholesale Trade)
15	CNS07	Num	Number of jobs in NAICS sector 44-45 (Retail Trade)
16	CNS08	Num	Number of jobs in NAICS sector 48-49 (Transportation and Warehousing)
17	CNS09	Num	Number of jobs in NAICS sector 51 (Information)
18	CNS10	Num	Number of jobs in NAICS sector 52 (Finance and Insurance)
19	CNS11	Num	Number of jobs in NAICS sector 53 (Real Estate and Rental and Leasing)
20	CNS12	Num	Number of jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)

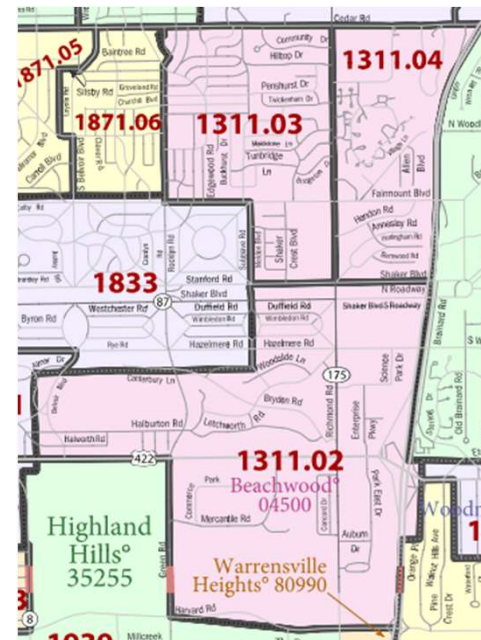
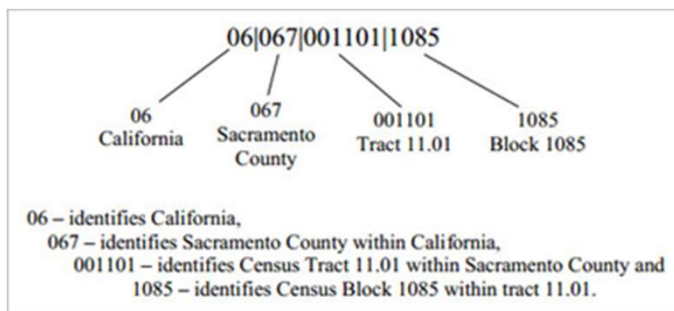
RAC: data of residents in Block Group

- Age, income levels, education levels, race, ethnicity, industry of work, etc.

WAC: data for workers in Block Group

Geographical Units in Census Bureau Data

Geocode Encoding Units:
Census Blocks
Census Tract
County
State





Big Data Preprocessing

Land Data

- Each file had 300+ variables, reduced to 190 relevant ones.
- Land Use data – Land Use Codes (LUC)
- Data Cleaning
- Data Integration

People Data

- Variables overlapped with supermarket access study
- 1,500 CSV files
- Normalized the data

Additional Geospatial Data

- Several Discretized variables
- Importance Measures
- Supermarket Access Information from Study's ArcGIS site



Integration of Geospatial Information

3 datasets integrated:

Integration of Different Encoding Scheme

- People Data used geocode, Land data used GPS for each Property

Integration of Tract information and Land Use Code (LUC)

Deriving Geospatial Information

Predictive Analysis Index (PAI) from Hot Spot Analysis

- Indicates how many properties per each Land use type (LUC) existed in each tract

Integration of County Planning data combined with LUC PAI analysis file and preprocessed people data

- Normalization



4

Framework Overview

Architecture & 4 Phases



Architecture of Food Desert Identifier System

1. Big Data Collection Phase

Gathered different data sources in varying formats

2. Big Data Preprocessing Phase

Data Transformation
Methods to reach a state that ML algorithms could process:

- Discretization
- Normalization

3. Deriving Geospatial Information

Using Data Analytic Methods:

- Hot Spot Analysis
- Predictive Accuracy Index



Architecture of Food Desert Identifier System

4. Feature Selection Phase

Process of reducing # of input variables when developing a predictive model

- Training Set
- Test Set

5. Classification

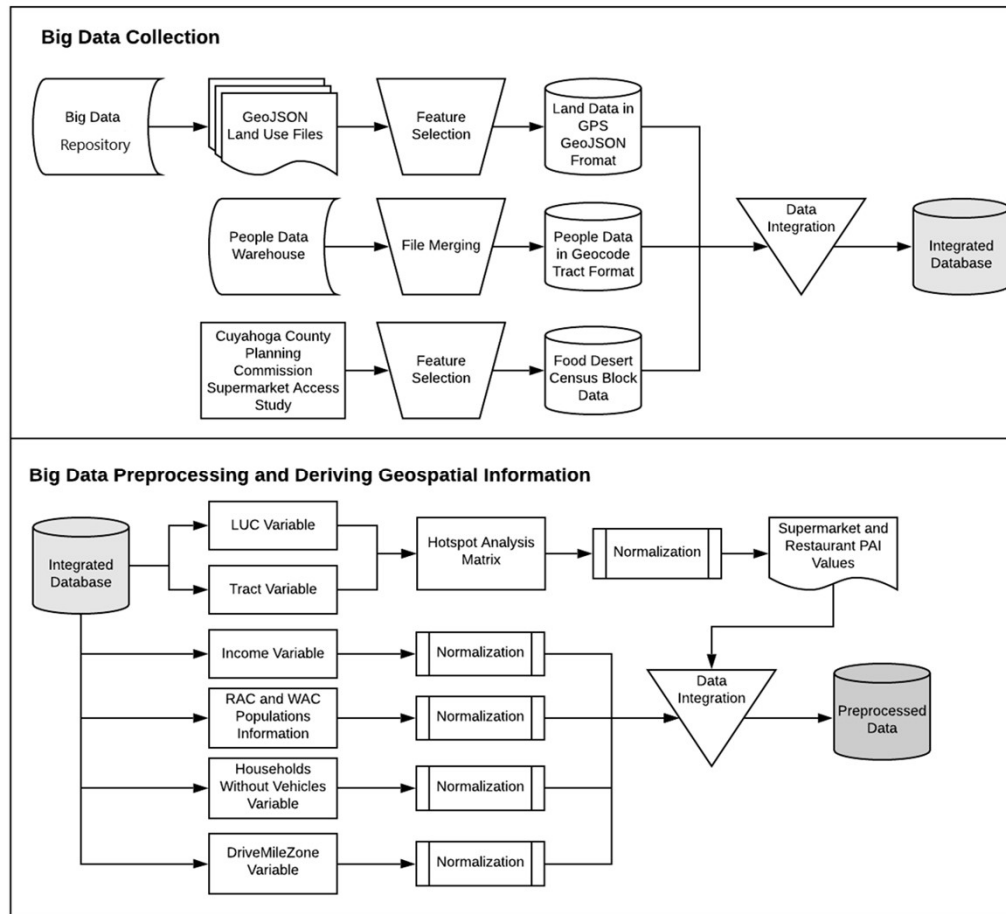
Machine Learning Algorithms:

- Random Forest
- Neural Network

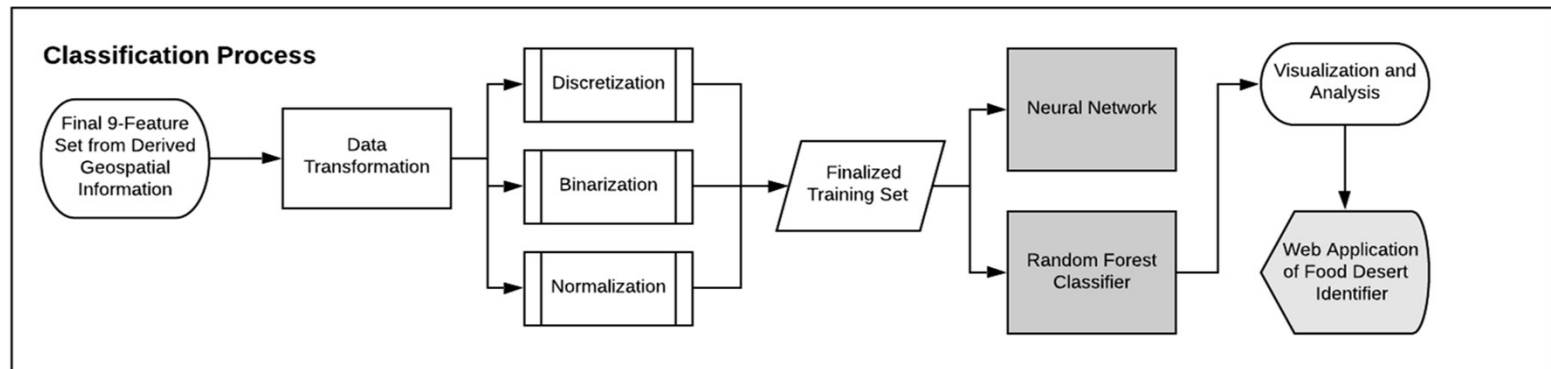
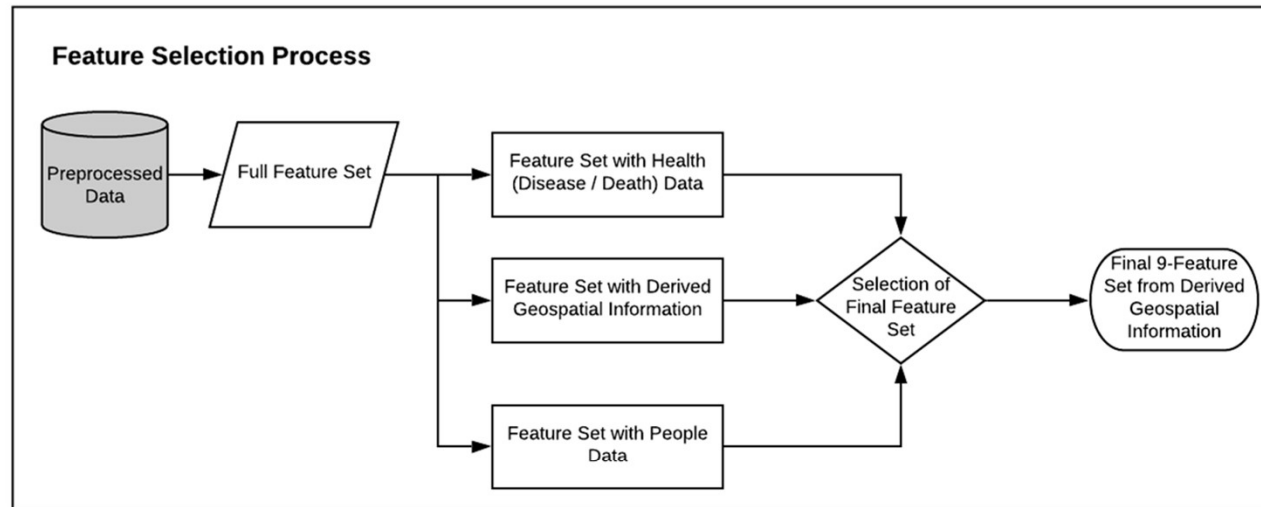
6. Visualization as Web Application

Website for User Interface for Food Desert Identifier

- Display on Map



Framework Overview for Food Desert Identifier System



Framework Overview for Food Desert Identifier System

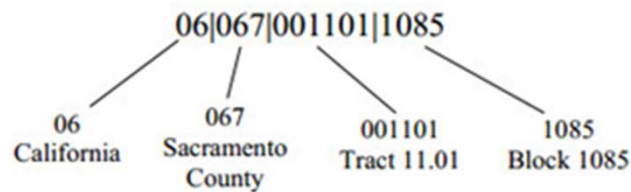
5

Methodology for Processing Geospatial Information

Geospatial information, hotspot
analysis, define feature
specifications



Geospatial Information Processing



06 – identifies California,
067 – identifies Sacramento County within California,
001101 – identifies Census Tract 11.01 within Sacramento County and
1085 – identifies Census Block 1085 within tract 11.01.

- Geocode – structured encoding scheme
- Census Block – last 4 digits, smallest unit of geography, tracts are made up of blocks



Method to Derive Geospatial Information: Hotspot Analysis

Hotspot Analysis

Process of identifying statistically significant locations (hot or cold spots)

- Converges points of proximity to one another based on distance, aggregates points of occurrence

$$\text{Prediction Accuracy Index (PAI)} = \frac{\frac{n}{N}}{\frac{a}{A}}$$

Prediction Accuracy Index (PAI)

Method of calculating & finding hot spots

- Hotspot PAI analysis:
Magnitude = indication of frequency



Defining Feature Specifications

9 Features

Residential areas, # of supermarkets, income level, households without vehicles, African American population, age groups, urban vs. suburban areas, # of restaurants, and DriveMileZone

6

Classification with Machine Learning

Selection of Features, Machine
Learning Classifiers



Selection of Features

- Originally 15 factors
 - Education, income levels
 - Demographics, transportation
 - Grocery store location, # of food stores
 - Urban vs. Suburban, # of children & seniors
 - Poverty measure, unemployment rate
 - Total housing units occupancy, % of population in group quarters
 - Low vehicle access tract, disease rates



Feature Set Categories

- Original 9 feature set
- People data feature set: 9 features & additional people data from RAC & WAC data
- Health-related disease feature set: 9 features + stroke, heart disease, cancer, and diabetes rates
- Derived geospatial feature set: 9 features + PAI values on supermarket & restaurant densities



Machine Learning Classifiers

Decision Tree

- Building block of random forest model
- Several individual decision trees operating as ensemble make up a Random Forest

Random Forest

- Ensemble Method: Wisdom of Crowds
- Trees operate as committee that outperforms individual models
- Features have predictive power, trees of forest are uncorrelated

Artificial Neural Network

- Multilayer Perceptron (MLP)
- Trained on set of input & output pairs
- Accuracy did not improve above 90%

7

Experiments

Evaluation of Classification and
Feature Sets



Evaluation of Classification

Randomness of Random Forest

- Variance reduction – combine diverse trees
- Fewer prediction errors
- Average predictions

Class Imbalance Problem

- Binary classification
- # of food desert negative class > # of food desert positive class
- Undersample normal and oversample positive

Accuracy of Classification

- Accuracy = $(TP + TN) / (TP + TN + FP + FN)$
- Recall = $TP / (TP + FN)$
- Precision = $TP / (TP + FP)$



Confusion Matrix with 9 Feature Set

Accuracy = 98%

Recall = 96%

Precision = 97%

	Predicted Class		
Actual Class		Class = Yes	Class = No
	Class = Yes	48523	1698
	Class = No	1697	117835



Confusion Matrix with 9 Feature Set + Derived Geospatial Information

Accuracy = 100%

Recall = 100%

Precision = 100%

	Predicted Class		
Actual Class		Class = Yes	Class = No
Class = Yes		48523	0
Class = No		0	117835

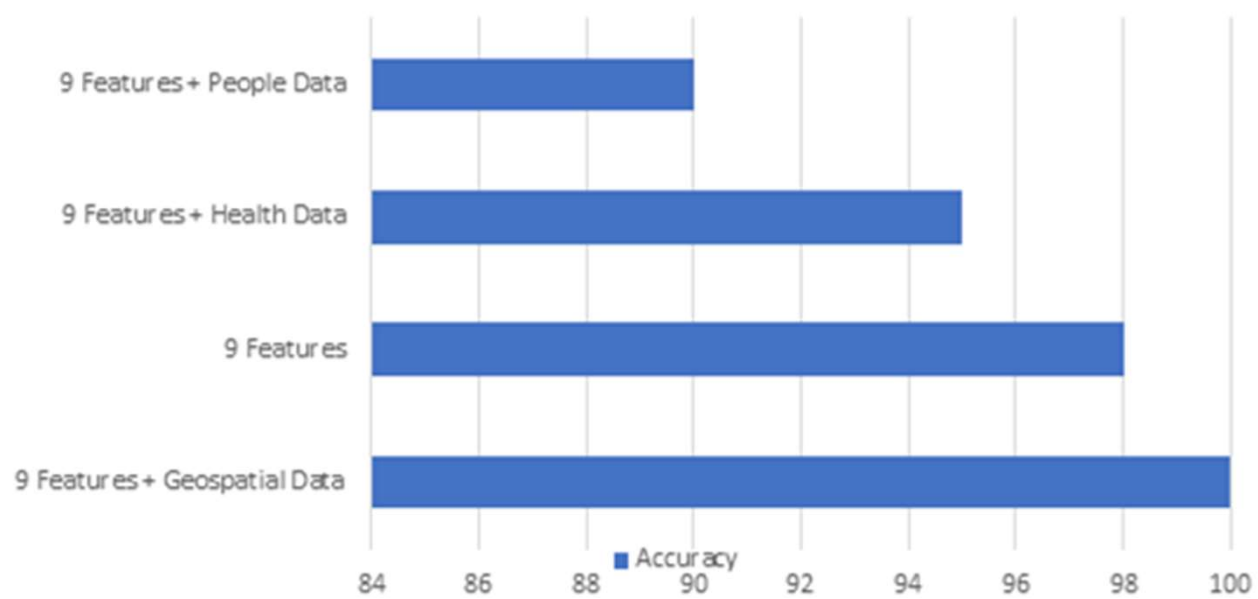


Discussion of Results

- Experiment Result indicates effectiveness of derived geospatial features
 - Accuracy: due to labeling training set with food desert census block data that considers 2 factors – the number of & distance to supermarkets.
- Limitation of the experiment due to lack of data of food deserts.
- The food desert census block was only data available that could've been used to label training set for evaluation



Accuracy





Discussion of Results

- Analysis of Different Accuracy of Feature Sets
 - 9 features + health data worse than 9 features, due to complex health data
 - Not always good indicator of poverty
 - 9 features + derived geospatial information has best result, due to the features like PAI from hotspot analysis
 - PAI and Distance to nearest supermarkets are effective in identifying food deserts



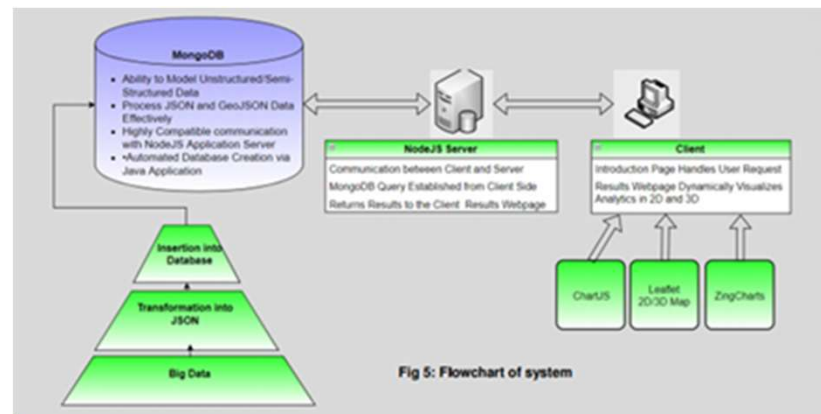
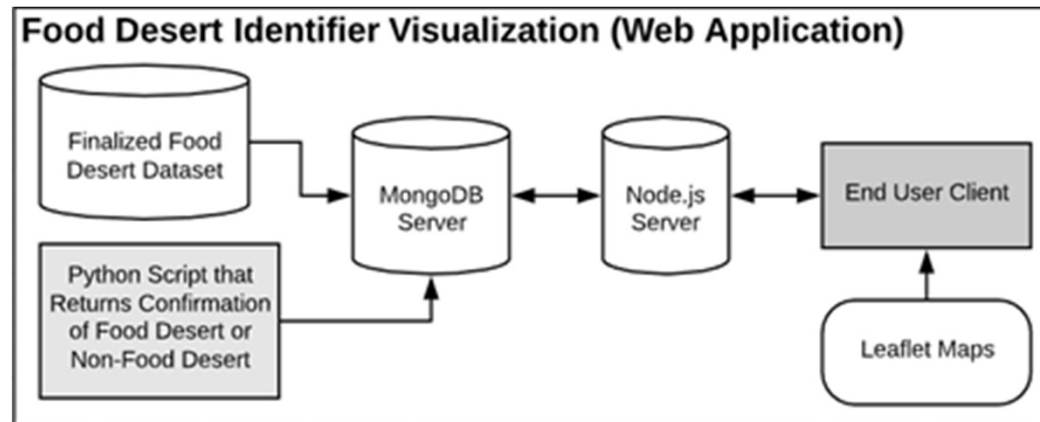
Discussion: Factors that Affect Accuracy

- Data Size, Class Distribution
- Complexity of Features
- Imbalance Class Distribution in data set
 - 48,523 values in Pos class for Food Desert areas
 - 117,835 values in Neg class for Non Food desert areas

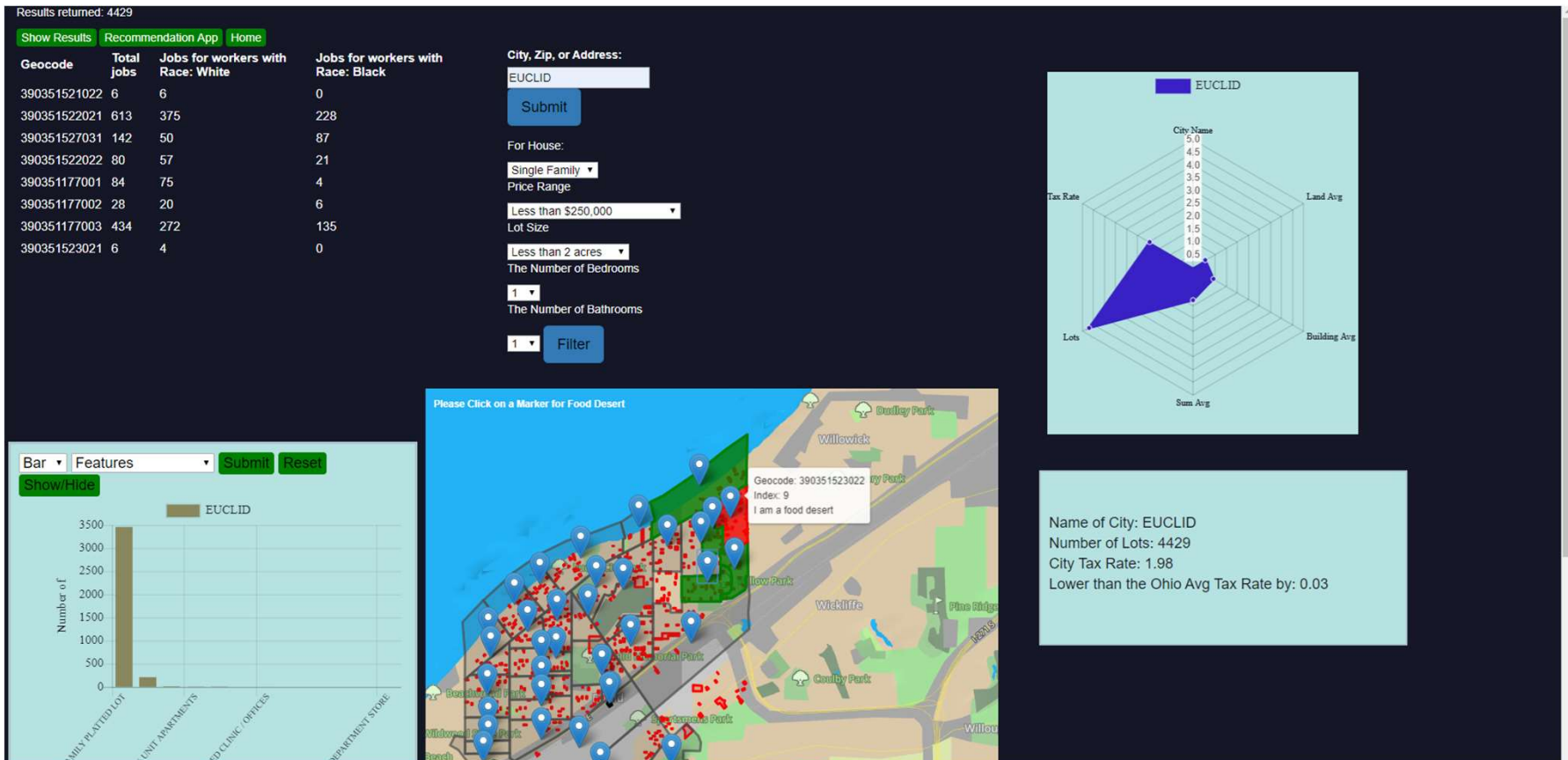
8

Web Application

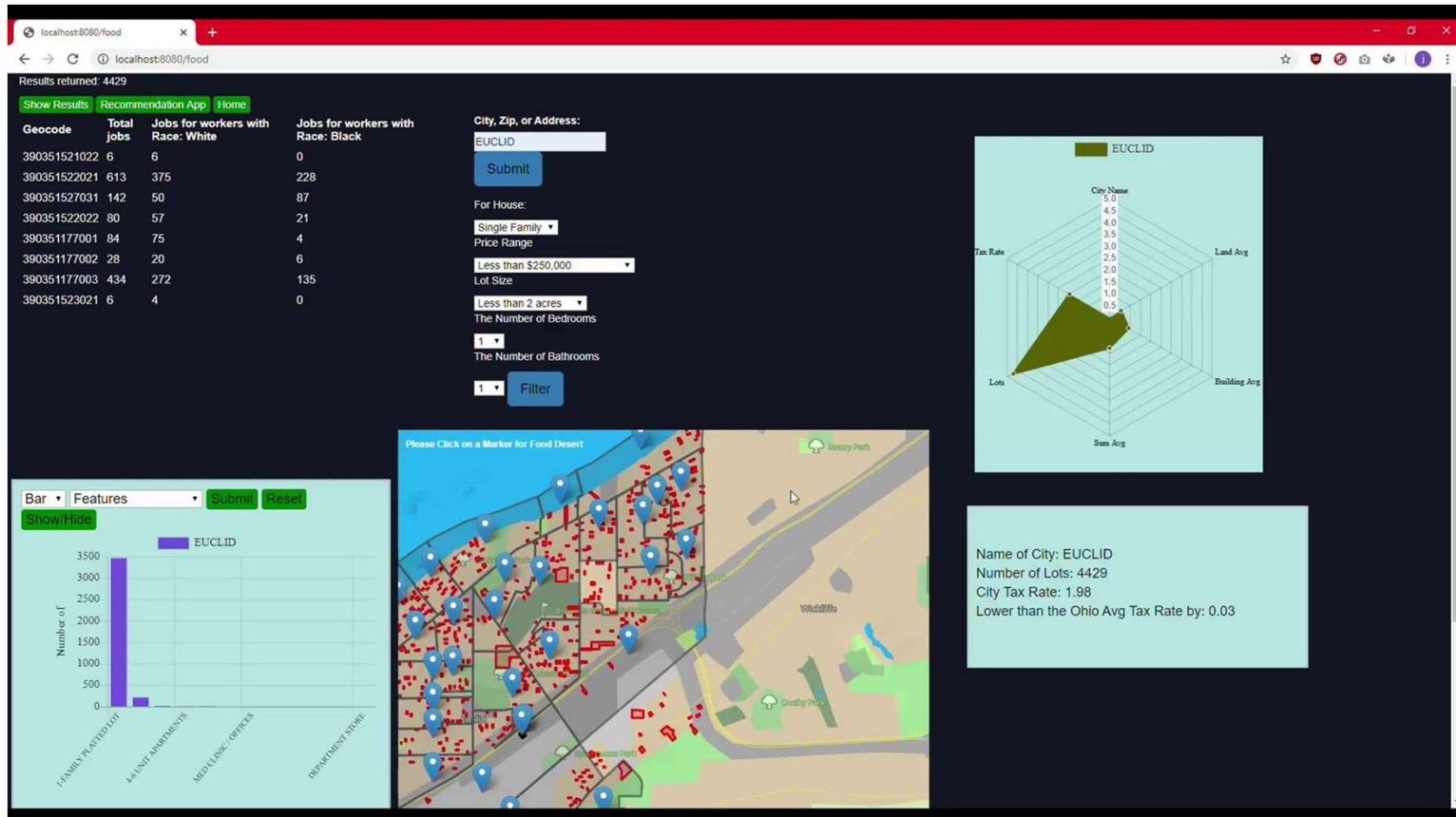
Food Desert Identifier Web App
Architecture, UI, Video Demo



Architecture of Data Flow of Web Application



Food Desert Identifier Web Example



Video Demo

Block (Geocode)	Food Desert (Y/N)
390351138011002	You are in a food desert
390351011011003	You are in a food desert
390351174002007	You are in a food desert
390351011011000	You are NOT in a food desert
390351011011001	You are NOT in a food desert
390351011021001	You are NOT in a food desert
390351035001022	You are NOT in a food desert

Some Identified Blocks of Food Deserts & Non-Food Deserts



9

Conclusion

Limitations



In Conclusion...

- Using big data analytics & ML classifiers, the food desert identifier locates low food access areas in Cuyahoga County
- A chance to improve awareness of food desert areas, to isolate factors that contribute to low food access



Limitations

- Lack of usable information & data for food deserts
 - Locating correct dataset (of which blocks are actual food deserts)
- Dependent on surveyed data generated from limited/inaccurate real world data



10

Future Work

Recommendations, Next Steps



For the future...

- Expand Food Desert Identifier to include other counties
- Select additional features, attach weights to components of food desert
- Create additional documentation
- Identify food mirages, swamps, oases

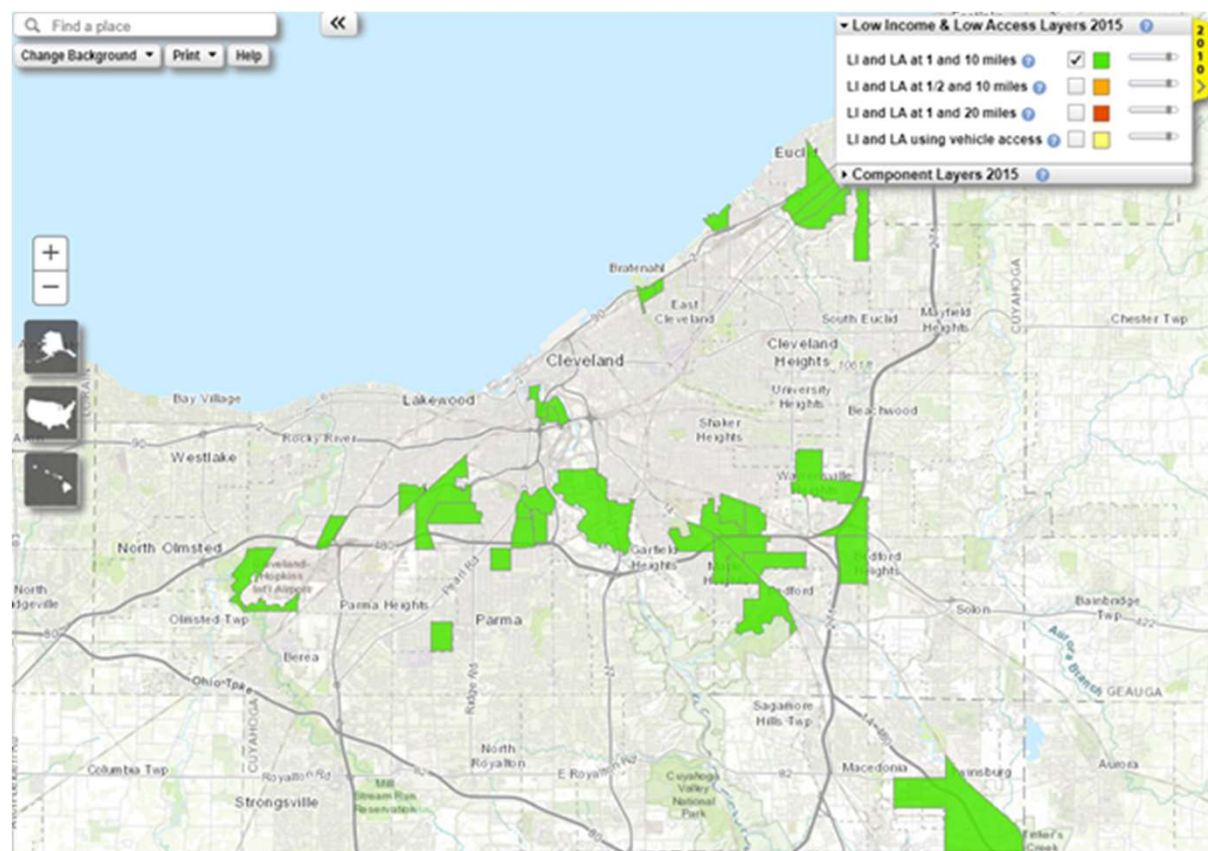
References

- [1] P. Dutko, M. V. Ploeg and T. Farrigan, "Characteristics and Influential Factors of Food Deserts," U.S. Department of Agriculture, Economic Research Service, 2012.
- [2] P. R. Rose, "Health Disparities, Diversity, and Inclusion," in *The Extent of the Health Status Gap and Why It Has Not Been Closed*, Jones & Bartlett Learning, 2018, pp. 23–25.
- [3] M. Barnes, "Solving the Problem of Childhood Obesity Within a Generation," White House Task Force on Childhood Obesity, 2010.
- [4] Feeding America, "Food Insecurity in Cuyahoga County," 2017. [Online]. [Accessed 1 April 2020].
- [5] Health Cleveland, "Active Living Map," [Online]. Available: <https://www.healthycle.org/active-living-maps>. [Accessed 1 May 2020].
- [6] J. New, "Center For Data Innovation," 27 October 2015. [Online]. [Accessed 13 April 2020].
- [7] S. A. Flocke, P. Ohri-Vachispati, E.-J. Shon, E. S. Trapl, E. Borawski, K. Matlack and D. A. Freedman, "Developing multidimensional measures of healthy food access among low-income adults in Cleveland, Ohio, USA," *Public Health Nutrition*, 2017.
- [8] T. Dubowitz, S. N. Zenk, B. Ghosh-Dastidar, D. A. Cohen, R. Beckman, G. Hunter, E. D. Steiner and R. L. Collins, "Health food access for urban food desert residents: examination of the food environment, food purchasing practices, diet and BMI," *Public Health Nutrition*, 2014.
- [9] A. Rhone, "Food Access Research Atlas," 31 October 2019. [Online]. [Accessed 13 April 2020].
- [10] M. Ver Ploeg, D. Nulph and R. Williams, "Mapping Food Deserts in the United States," 1 December 2011. [Online]. [Accessed 13 April 2020].
- [11] Cuyahoga County Planning Commission, "Cuyahoga County Supermarket Assessment: 2018 Inventory Update," Cleveland, 2019.
- [12] M. S. LeClair and A.-M. Aksan, "Redefining the food desert: combining GIS with direct observation to measure food access," *Agriculture and Human Values*, 15 January 2014.
- [13] V. P. Sisipiku and N. Barbour, "Use of GIS Spatial Analysis to Identify Food Deserts in the State of Alabama," *Athens Journal of Health*, pp. 91–104, 2014.
- [14] M. Acconciamezza, D. Iqbal and B. Putman, "Food Desert Predictor," 2017. [Online]. Available: <https://www.ischool.berkeley.edu/projects/2017/food-desert-predictor>. [Accessed 31 March 2020].
- [15] E. Major, E. C. Delmelle and E. Delmelle, "SNAPScares: Using Geodemographic Segmentation to Classify the Food Access Landscape," *Urban Science*, pp. 1–20, 2018.
- [16] S. Chainey, L. Thompson and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," *Security*, vol. 21, pp. 4–28, 2008.
- [17] NODIS, "Northern Ohio Data and Information Service (NODIS) Ohio Tracts & Blockgroups," [Online]. [Accessed 13 April 2020].
- [18] Cuyahoga County GIS, "Cuyahoga County Open Data," 2017. [Online]. Available: <https://data-cuyahoga.opendata.arcgis.com/>. [Accessed 13 April 2020].
- [19] GeoJSON.org, "geojson.org," [Online]. Available: geojson.org. [Accessed 13 April 2020].
- [20] Cuyahoga County GIS, "Cuyahoga County Department of Information Technology GIS Data," [Online]. Available: <http://gis.cuyahogacounty.us/en-US/GIS-Data.aspx>. [Accessed 13 April 2020].

- [21] Cuyahoga County Fiscal Officer, "Real Property Informaiton," [Online]. Available: <https://fiscalofficer.cuyahogacounty.us/en-US/REPI.aspx>. [Accessed 13 April 2020].
- [22] United States Census Bureau, "LEHD Origin-Destination Employment Statistics (LODES)," [Online]. [Accessed 13 April 2020].
- [23] USC GIS Blog, "How GIS Helps to Overcome Urban Food Deserts," 14 June 2019. [Online]. Available: <https://gis.usc.edu/blog/overcoming-urban-food-deserts/>. [Accessed 13 April 2020].
- [24] United States Census Bureau, "Understanding Geographic Identifiers (GEOIDs)," 1 April 2020. [Online]. [Accessed 9 April 2020].
- [25] I. Lendel, O. Pasha, D. Sohrabian and M. Ellerbrock, "GCRTA Economic Impact and Contributions to Local Economy," 5 March 2019. [Online]. Available: http://www.riderta.com/sites/default/files/GCRTA_economic_impact_and_contributions_to_local_economy.pdf. [Accessed 1 May 2020].
- [26] J. Hunt, "Do Crime Hot Spots Move? Exploring the Effects of the Modifiable Areal Unit Problem and Modifiable Temporal Unit Problem on Crime Hot Spot Stability," 2016.
- [27] H. Fang, Z. Zhang, C. J. Wang, M. Daneshmand, C. Wang and H. Wang, "A survey of big data research," IEEE network, pp. 6-9, 2015.
- [28] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Benitez and F. Herrera, "Big data preprocessing: methods and prospects," Big Data Analytics, 2016.
- [29] R. Shaikh, "Feature Selection Techniques in Machine Learning with Python," 28 October 2018. [Online]. [Accessed 13 April 2020].
- [30] Federal Communication Commission, "More About Census Blocks," 2015.
- [31] United States Census Bureau, "Census Tracts and Block Numbering Areas," pp. 1-17.
- [32] Geospatiality, "What is Hotspot Analysis?," 21 January 2016. [Online]. Available: <https://glenbambrick.com/2016/01/21/what-is-hotspot-analysis/>. [Accessed 13 April 2020].
- [33] J. E. Eck, S. Chainey, J. G. Cameron, M. Leitner and R. E. Wilson, "Mapping Crime: Understanding Hot Spots," National Institute of Justice, 2005.
- [34] Census.gov, "Poverty Thresholds," 22 January 2020. [Online]. Available: <https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html>. [Accessed 14 April 2020].
- [35] J. Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2011.
- [36] G. Louppe, "Understanding Random Forests: From Theory to Practice," 2014.
- [37] P.-n. Tang, M. Steinbach, V. Kumar and A. Wesley, Introduction to Data Mining.
- [38] UFLDL Stanford Tutorial, "Multi-Layer Neural Network," [Online]. Available: <http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/>. [Accessed 17 April 2020].
- [39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, pp. 2825-2830, 2011.
- [40] Tan, Steinbach and Kumar, "Data Mining Classification: Basic Concepts, Decision Trees, and Model Evaluation," [Online]. Available: http://eecs.csuohio.edu/~sschung/CIS593/CIS593_Chap4_basic_classification_DTIntro.pdf. [Accessed 13 April 2020].
- [41] S. Cummins and S. Macintyre, "'Food desert' -- evidence and assumption in health policy making," BMJ (Clinical research ed.), 2002.

THANK YOU!

Any questions?

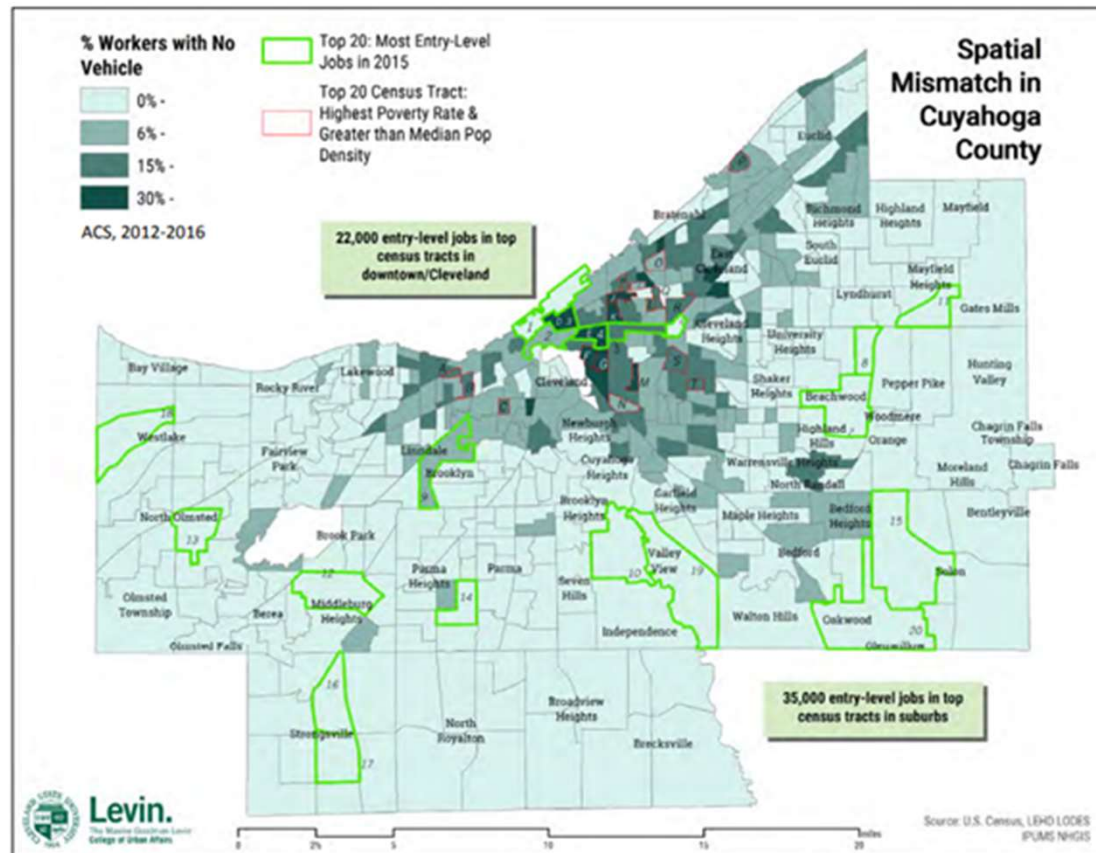


Food Access Research Atlas with One Layer Applied

Join_Count	TARGET_FID	GEOID10 ▲	AreaSqMi	SubSPA
1	15,049	39035101101100	0.01	Cleve: Cudell
1	7,385	39035101101100	0.01	Cleve: Cudell
1	12,274	39035101101100	0.01	Cleve: Cudell
1	6,567	39035101101200	0.01	Cleve: Cudell
1	579	39035101101200	0.00	Cleve: Cudell
1	2,581	39035101101200	0.02	Cleve: Cudell
1	12,001	39035101101200	0.01	Cleve: Cudell

Population at 200% and over of Poverty Level	Population Below Poverty Level	Population - African American	Meets Poverty Threshold (30% of Pop Below 200% Poverty)
20.72	85.77	47.19	Y
20.01	82.81	45.57	Y
35.61	147.38	81.10	Y
26.36	77.64	88.91	Y
8.53	25.12	28.76	Y
68.08	200.48	229.59	Y

Partial Documentation of Data from Supermarket Access Study



Distribution of Area with Low Vehicle Ownership

LUC COUNTS			TRACT		
LUC_Meaning	LUC	Count	101101	101102	101200
Commercial campgrounds	4160	2	0	0	0
Other commercial housing	4190	20	0	0	0
Small (under 7500 sq. ft.) detached retail	4200	1693	0	0	0
Supermarkets	4210	90	0	0	0
Discount stores and junior department stores	4220	165	0	0	0
Furniture marts	4230	29	0	0	0
Full line department store	4240	35	0	0	0
Strip center retail (4 or more, > 7500 sq. ft.)	4250	449	0	0	0
NULL (Missing LUC meaning)	4255	35	0	0	0
Community shopping center	4260	110	0	0	0
Regional shopping center	4270	70	0	0	0
Franchise food stores	4280	68	0	0	0
Other retail structures	4290	425	0	0	0
NULL (Missing LUC meaning)	4295	24	0	0	0
Restaurant, cafeteria	4300	307	0	0	0
Nightclub restaurant	4301	68	0	0	0
Party center	4302	37	0	0	0

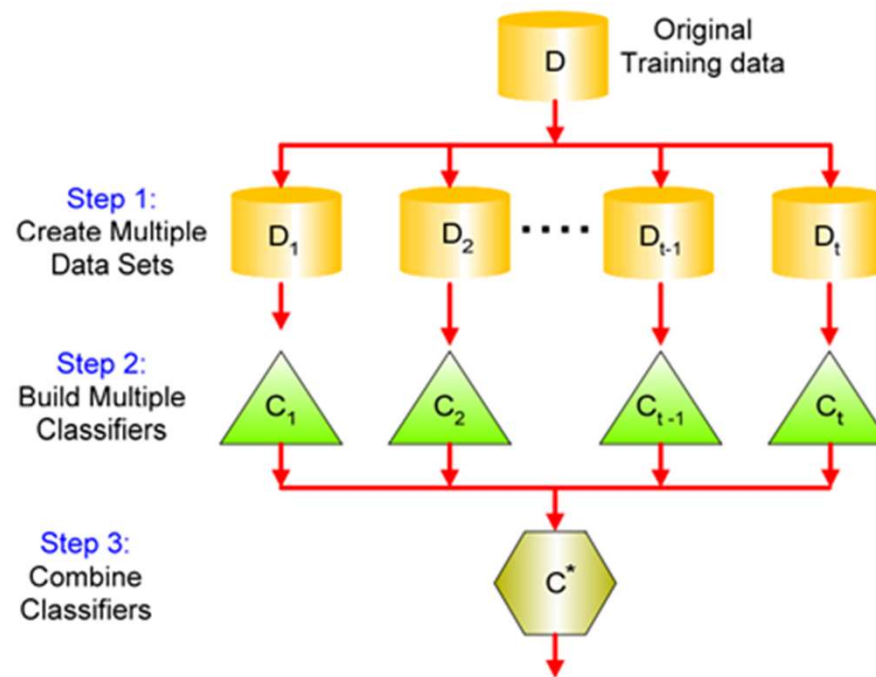
A Part of the LUC PAI Analysis Matrix


$$GINI(t) = 1 - \sum_j [p(j|t)]^2$$

$$GINI_{split} = \sum_{i=1}^k \frac{n_i}{n} GINI(i)$$

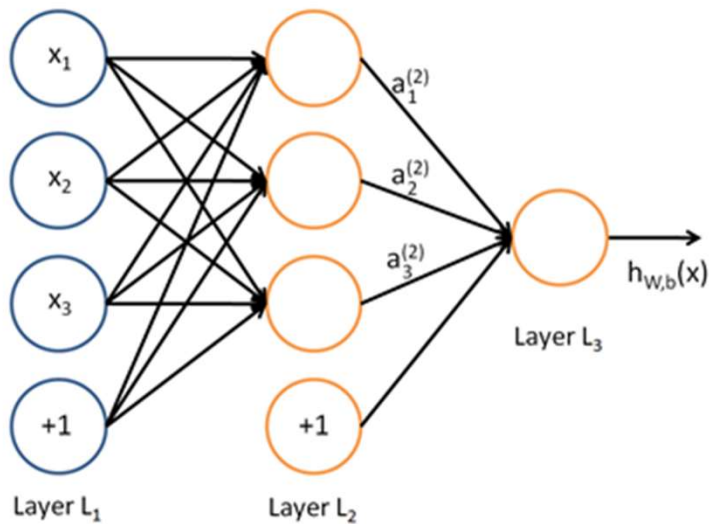
$$Entropy(t) = -\sum_j p(j|t) \log p(j|t)$$

Decision Tree Equations



$$C^*(x) = \operatorname{argmax}_y \sum_{j=1}^T \delta(C_j(x) = y)$$

Ensemble Method Example & Equation



$$\begin{aligned}
 a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}) \\
 a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}) \\
 a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}) \\
 h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})
 \end{aligned}$$

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

$$\begin{aligned}
 z^{(2)} &= W^{(1)}x + b^{(1)} \\
 a^{(2)} &= f(z^{(2)}) \\
 z^{(3)} &= W^{(2)}a^{(2)} + b^{(2)} \\
 h_{W,b}(x) = a^{(3)} &= f(z^{(3)})
 \end{aligned}$$

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{W,b}(x^i) - y^i\|^2 \right) \right]$$

Neural Network Example & Equations