

Food Desert Prediction Strategy

Mathematical Formulations and Methodology

Research Question:

Can we predict which U.S. census tracts are at risk of becoming food deserts in the next decade, based on demographic shifts, retailer patterns, transit access, and socioeconomic change?

Feature Normalization

Min-Max Normalization:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Z-Score Normalization:

$$z = \frac{x - \mu}{\sigma}$$

Population Normalized Features:

$$\text{normalized_lapop} = \frac{\text{lapop1}}{\text{Pop2010}}$$

Low-Income Normalized:

$$\text{normalized_lalowi} = \frac{\text{lalowi1}}{\text{TractLOWI}}$$

Income Score:

$$\text{income_score} = \frac{\text{MedianIncome} - \min_{\text{income}}}{\max_{\text{income}} - \min_{\text{income}}}$$

Poverty Score:

$$\text{poverty_score} = \frac{\text{PovertyRate}}{100}$$

Education Score:

$$\text{education_score} = \frac{\text{BachelorDegreeCount}}{\text{TotalPopulation}}$$

Rent Burden:

$$\text{rent_burden} = \frac{\text{MedianGrossRent}}{\text{MedianFamilyIncome}}$$

Category Scores and Risk Calculation

Category Score:

$$\text{Category_Score} = \sum_i (w_i \times f_i)$$

Feature Category Weights:

$w_A = 0.25$ (Food Access Status)

$w_B = 0.30$ (Socioeconomic Indicators)

$w_C = 0.25$ (Retail Environment Changes)

$w_D = 0.15$ (Transportation Access)

$w_E = 0.05$ (Demographic and Health Indicators)

Overall Risk Score:

$$\text{Risk_Score} = w_A \times A + w_B \times B + w_C \times C + w_D \times D + w_E \times E$$

$$\text{Risk_Score} = 0.25 \times A + 0.30 \times B + 0.25 \times C + 0.15 \times D + 0.05 \times E$$

Probability Conversion (Logistic Function):

$$P = \frac{1}{1 + e^{-(\text{Risk_Score} - \theta)}}$$

where θ is the decision threshold

Social Vulnerability Index (SVI) Calculation

SVI Score Formula:

$$SVI = 0.30 \times S_{\text{socio}} + 0.25 \times S_{\text{household}} + 0.25 \times S_{\text{minority}} + 0.20 \times S_{\text{housing}}$$

Socioeconomic Status (30%):

$$S_{\text{socio}} = 0.40 \times \text{Poverty} + 0.30 \times \text{Income} + 0.20 \times \text{Education} + 0.10 \times \text{Rent}$$

Household Composition (25%):

$$S_{\text{household}} = 0.50 \times \text{LowIncome} + 0.50 \times (1 - \text{VehicleOwnership})$$

Minority Status (25%):

$$S_{\text{minority}} = 0.60 \times \text{LowIncomeProxy} + 0.40 \times \text{Poverty}$$

Housing & Transportation (20%):

$$S_{\text{housing}} = 0.50 \times (1 - \text{VehicleOwnership}) + 0.30 \times \text{Crowding} + 0.20 \times \text{Rent}$$

Interpretation:

0.0 – 0.3: Low vulnerability

0.3 – 0.5: Moderate vulnerability

0.5 – 0.7: High vulnerability

0.7 – 1.0: Very high vulnerability

Additional Feature Calculations

Grocery Store Density Change:

$$\Delta_{\text{density}} = \frac{\text{stores}_t - \text{stores}_{t-5}}{\text{stores}_{t-5}}$$

Density Risk Score:

$$\text{risk_score} = -\Delta_{\text{density}}$$

Vehicle Ownership Rate:

$$\text{vehicle_ownership} = 1 - \frac{\text{HouseholdsNoVehicle}}{\text{TotalHouseholds}}$$

Transit Access Score:

$$\text{transit_access} = \text{TransitFrequency} \text{ if accessible, else 0}$$

Income-Based Vulnerability:

$$\text{income_vuln} = 1.0 \text{ if income} < \$30K, 0.7 \text{ if } \$30K-50K, 0.4 \text{ if } \$50K-75K, 0.1 \text{ if } \geq \$75K$$

Rent Burden Score:

$$\text{rent_burden} = 1.0 \text{ if rent} > 50\% \text{ income}, 0.7 \text{ if } 30-50\%, 0.3 \text{ if } \leq 30\%$$

Model Performance Metrics

ROC-AUC:

$$AUC = \int_0^1 TPR(FPR^{-1}(x))dx$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Risk Level Classification:

High Risk: $P \in [0.70, 1.00]$

Moderate Risk: $P \in [0.40, 0.69]$

Low Risk: $P \in [0.20, 0.39]$

Very Low Risk: $P \in [0.00, 0.19]$