DEVFEST TROY 2026 · RENSSELAER POLYTECHNIC INSTITUTE

Democratizing Civic Data: Transforming Public Datasets into Policy

A blueprint for cleaning messy civic datasets, extracting grassroots priorities, and driving municipal decisions.

Neethu Thalappan Koroth | Enterprise Asset Management Analyst, MTA · NYIT CS

THE CIVIC DATA GAP

Open Data Portals Fail to Reach Policy Desks

Raw Portals Are Inaccessible

Millions of records live in unindexed formats without narrative context.

Leaders Require Decisive Evidence

Municipal leaders lack bandwidth to write custom SQL scripts.

FIELD VALIDATION

The Whole Health Datathon Sprint

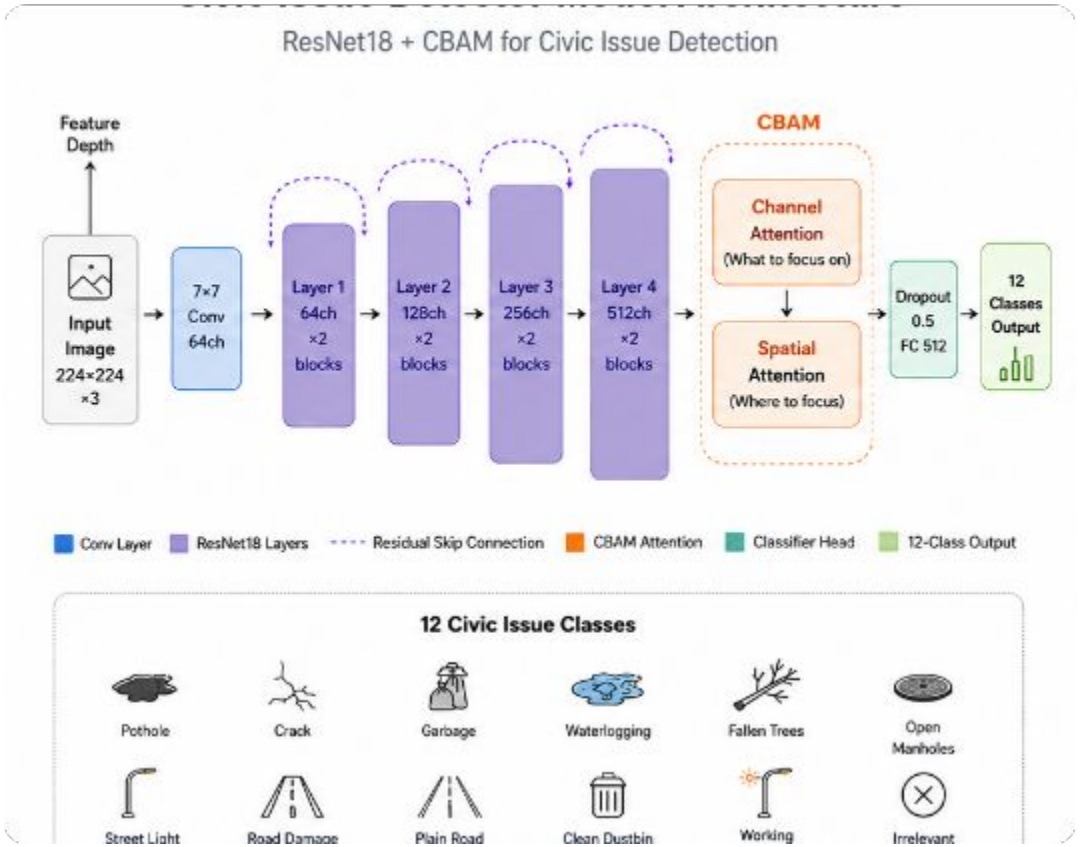
48h
Rapid analysis sprint timeline.

$5M
NYC civic participatory funding.

#1
Awarded Best Policy Brief.

CONTEXTUAL BLUEPRINT

From Messy Geographic Rows to Targeted Civic Impact



1. Spatial Inconsistency

Disparate community district boundaries across five boroughs.

2. Multi-Lingual Free Text

Thousands of unnormalized citizen submissions in 12 languages.

ENGINEERING ARCHITECTURE

The 4-Stage Data Transformation Workflow

STAGE 01

Ingestion & Schema

Automated schema unification and validation.

STAGE 02

NLP & Clustering

Semantic clustering of open citizen proposals.

STAGE 03

Geospatial Overlay

Correlating community needs with health indices.

STAGE 04

Policy Translation

Synthesizing data into 1-page executive memos.

EMPIRICAL FINDINGS

What Communities Actually Voted For



ENGINEERING RIGOR

Overcoming Public Data Complexities

Free-Text NLP Clustering

TF-IDF and sentence embeddings group unstructured citizen proposals into discrete categories.

`Result: Zero manual tagging overhead.`

Geospatial Join Accuracy

Bounding polygons normalize neighborhood voting tracts against health disparity census areas.

`Result: High-confidence spatial attribution.`

POLICY TRANSLATION

Structure of a Winning 1-Page Policy Brief

The Evidence Hook

One bold statistical takeaway.

Demographic Context

Under-served population impact metrics.

Actionable Roadmap

Targeted budget allocation breakdown.

CIVIC TURNOUT REALITY

The Power of Hybrid Voting Channels

77%

Ballots submitted via physical paper at community partner centers.

LOCAL IMPACT

Replicating the Framework in Upstate New York

Municipal Portals

Partner directly with Troy and Albany county offices.

University Pipelines

Engage RPI and regional CS students in civic hackathons.

Open Data Standard

Publish open-source pipeline scripts for smaller townships.

NEXT STEPS

How You Can Build for Public Good

- **Identify:** Audit regional open data portals for unaddressed public needs.
- **Clean:** Write automated extract-transform scripts and document discrepancies.
- **Deliver:** Package conclusions into 1-page briefs tailored for council members.

THANK YOU

Let’s Democratize Civic Tech Together

Questions, feedback, or collaborations on open municipal pipelines?

linkedin.com/in/nithu-nair neethunairkoroth@gmail.com

Image Sources



https://miro.medium.com/1*ZdTZ__A3_11Q5ZFzhw3geA.png

Source: medium.com