Idea #3: Predicting Opioid Overdose Risk at Community Level

Acceptance Likelihood: 4.3/5 — Strong Pick

The Problem

Canada's opioid crisis is one of the most devastating public health emergencies in the country's history, driven by the proliferation of illicitly manufactured fentanyl.

Key Statistics

- **40,000+ opioid-related deaths** in Canada since 2016
- British Columbia is the hardest-hit province with the highest per-capita death rate
- **2023:** Over 8,000 apparent opioid toxicity deaths nationally
- Overdose rates surged during COVID-19 and have remained elevated
- The crisis disproportionately affects men aged 20-49, Indigenous populations, and those experiencing homelessness
- 86% of accidental opioid deaths involve fentanyl or fentanyl analogues

Why It Matters

Opioid overdoses are preventable deaths. Predicting which communities face the highest risk enables **targeted deployment** of harm reduction resources: naloxone distribution, supervised consumption sites, treatment programs, and outreach workers. Every correctly predicted hotspot can save lives.

Why Machine Learning Can Help

Overdose patterns are influenced by identifiable community-level factors:

- **Prescription opioid dispensing rates**
- **Socioeconomic deprivation** (unemployment, income, housing instability)
- **Mental health and addiction service availability**
- **Historical overdose rates and trends**
- **Demographic composition** (age, gender, Indigenous population)
- **Urban vs. rural classification**
- **Proximity to harm reduction services**

ML can learn complex interactions between these factors to predict community-level overdose risk.

ML Task

Regression/classification: Given community characteristics, predict overdose rate or risk category for the next quarter.

Published Approaches

- **Gradient boosting:** Achieved **83.7% balanced accuracy** for individual overdose prediction in Canadian province
- **Logistic regression:** 30-day risk prediction after opioid dispensation in Alberta

- **Cluster analysis:** StatCan used ML to identify distinct overdose patient profiles (intersectional approach)

Datasets

1. BC Opioid Overdose Analytical File (BCOOAF)

- **Source:** Statistics Canada (linked data)
- **URL:** https://www150.statcan.gc.ca/n1/pub/82-003-x/2023003/article/00001-eng.htm
- **Contains:** 13,318 opioid overdose records (2014-2016) linked to health, employment, social assistance, and police contact data
- **Note:** May require research access through StatCan Research Data Centre

2. Federal/Provincial Opioid Surveillance Data

- **Source:** Public Health Agency of Canada
- **Contains:** National and provincial overdose death counts, quarterly updates, demographic breakdowns
- **Format:** Public reports and downloadable tables
- **Use:** Outcome variable for community-level prediction

3. Canadian Community Health Survey (CCHS)

- **Source:** Statistics Canada
- **Contains:** Health behaviors, substance use, mental health, socioeconomic data at community level
- **Format:** Public use microdata files available
- **Use:** Feature engineering for community-level risk factors

---

Proposed ML Pipeline

```
1. Data collection
   - Aggregate provincial overdose surveillance data by health region
   - Download CCHS community health profiles
   - Merge with Census socioeconomic data (income, employment, housing)

2. Feature engineering
   - Community deprivation indices
   - Healthcare access scores (distance to treatment, harm reduction sites)
   - Historical overdose trend features (rolling averages, rate of change)
   - Demographic composition features

3. Model training
   - XGBoost regressor (primary — predict overdose rate)
   - Random Forest classifier (classify risk tiers: low/medium/high)
   - Logistic regression (interpretable baseline)

4. Evaluation
   - MAE/RMSE for regression, AUC-ROC for classification
   - Feature importance (SHAP values)
   - Geographic validation (do predictions match known hotspots?)

5. Deliverable
```

```
- Risk map of Canadian communities
- Dashboard with key risk factors per region
- Recommendations for resource allocation
```

Strengths

- **Extremely urgent:** Active crisis with thousands dying annually
- **Strong ML precedent:** 83.7% accuracy already demonstrated in Canadian data
- **High impact:** Predictions directly inform where to deploy naloxone, outreach, treatment
- **Novel for Let's SOLVE It:** Never attempted in past cohorts
- **StatCan data infrastructure:** BCOOAF and CCHS provide rich feature sets

Risks & Mitigations

Risk	Mitigation
Individual-level data restricted	Work at community/health-region level using aggregated public data
Sensitivity of topic	Frame as supporting harm reduction. Avoid stigmatizing language.
Data access delays	Use publicly available surveillance reports as primary. BCOOAF as stretch goal.
Model could stigmatize communities	Focus on resource allocation framing, not community labeling

---

References

- Molecular Psychiatry (2025): "Population-level individualized prospective prediction of opioid overdose using machine learning" — https://www.nature.com/articles/s41380-025-02992-4
- BMC Public Health (2021): "Safe opioid prescribing: a prognostic ML approach in Alberta, Canada" — https://pmc.ncbi.nlm.nih.gov/articles/PMC8160164/
- StatCan (2023): "Exploring the intersectionality of characteristics among those who experienced opioid overdoses" — https://www150.statcan.gc.ca/n1/pub/82-003-x/2023003/article/00001-eng.htm
- Lancet Digital Health (2021): "Big data and predictive modelling for the opioid crisis" — https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00058-3/fulltext