Idea #1: Predicting Drinking Water Advisories on First Nations Reserves

Acceptance Likelihood: 4.8/5 — Top Pick

The Problem

Indigenous communities in Canada face a persistent drinking water crisis. As of recent reporting, approximately 30 long-term drinking water advisories (DWAs) remain active on First Nations reserves — some lasting over a decade.

Key Statistics

- As of April 2016 baseline, **78 long-term DWAs** were identified for remediation
- Many communities have been under advisories for **10+ years**
- Saskatchewan and Ontario experience the most frequent and long-lasting DWAs
- First Nations reserves are **disproportionately affected** compared to non-Indigenous communities
- Water systems on reserves are often underfunded, understaffed, and aging

Why It Matters

Access to clean drinking water is a basic human right. The ongoing water crisis on First Nations reserves is one of Canada's most visible failures of infrastructure equity. Predicting which systems will fail allows **proactive intervention** — fixing problems before communities lose access to safe water.

Why Machine Learning Can Help

DWAs are not random. They follow predictable patterns based on system characteristics:

- **Source water type** (groundwater vs. surface water)
- **System age and infrastructure condition**
- **Operator certification level**
- **Population served**
- **Province/region** (geographic and climate factors)
- **Historical advisory frequency**

ML can learn these patterns to predict which water systems are at highest risk of issuing new advisories, enabling targeted infrastructure investment.

ML Task

Binary classification: Given a water system's characteristics, predict whether it will issue a DWA in the next 12 months.

Proven Approaches

- **XGBoost:** Published research achieved **86% accuracy** on this exact prediction task
- **Decision trees:** Earlier data mining study successfully mapped system characteristics to DWA likelihood
- **Feature importance:** Province location, system age, operator certification are top predictive features

---

Datasets

1. ISC Long-Term DWA Dataset (Primary)

- **Source:** Indigenous Services Canada via Open Canada Portal
- **URL:** https://open.canada.ca/data/en/dataset/57b86ac5-e127-41bc-94b8-14b2d89aed0b
- **Contains:** Progress tracking on 78 long-term DWAs, system details, advisory status, dates
- **Format:** Open data, downloadable

2. Canadian Environmental Sustainability Indicators (CESI)

- **Source:** Environment and Climate Change Canada
- **URL:** https://www.canada.ca/en/environment-climate-change/services/environmental-indicators.html
- **Contains:** Drinking water quality indicators, quality of life metrics
- **Use:** Supplementary environmental context for model features

3. StatCan Quality of Life — Drinking Water

- **Source:** Statistics Canada
- **URL:** https://www160.statcan.gc.ca/environment-environnement/drinking-water-eau-potable-eng.htm
- **Contains:** National and regional drinking water quality metrics
- **Use:** Baseline comparison between Indigenous and non-Indigenous water quality

---

Proposed ML Pipeline

```
1. Data collection & cleaning
   - Download ISC DWA dataset
   - Extract system characteristics (age, source, operator, population, province)
   - Engineer features (historical advisory count, duration patterns)

2. Exploratory analysis
   - Geographic distribution of DWAs
   - Correlation analysis between features and advisory status
   - Temporal trends in advisory issuance/resolution

3. Model training
   - XGBoost classifier (primary)
   - Random Forest (baseline comparison)
   - Logistic Regression (interpretable baseline)
   - Cross-validation with stratified splits

4. Evaluation
   - Accuracy, precision, recall, F1, AUC-ROC
   - Feature importance analysis (SHAP values)
   - Geographic analysis of model predictions

5. Deliverable
   - Interactive dashboard showing risk scores by community
   - Feature importance visualization
   - Recommendations for infrastructure prioritization
```

Strengths

- **Perfect equity alignment:** Directly addresses Indigenous rights — core to RBC Borealis + CIFAR diversity mission
- **Uniquely Canadian:** This problem exists nowhere else in the same form
- **Government data on Open Canada:** No access barriers
- **Published ML precedent:** 86% accuracy already demonstrated — validates feasibility
- **Actionable output:** Risk predictions can direct infrastructure funding
- **Never done in past cohorts:** Completely novel for Let's SOLVE It

Risks & Mitigations

Risk	Mitigation
Data sovereignty concerns	Acknowledge in proposal. Use only open government data. Frame as supporting Indigenous communities, not extracting from them.
Small dataset size	78 long-term DWAs is small — augment with shorter-term advisories and historical records
Sensitivity of topic	Frame carefully with respect. Focus on infrastructure systems, not communities as problems.
Scope creep	Keep focused on DWA prediction. Don't try to solve the entire water crisis.

References

- MDPI Water (2024): "Key Factors Influencing Drinking Water Advisories on Indigenous Reserves in Canada: An XGBoost Analysis" — https://www.mdpi.com/2073-4441/16/24/3647
- Springer (2015): "Using Data Mining to Understand Drinking Water Advisories in Small Water Systems" — https://link.springer.com/article/10.1007/s11269-015-1108-6
- StatCan SDG Indicator 6.1.1: Number of long-term DWAs on public systems on reserves — https://sdgcif-data-canada-oddcic-donnee.github.io/6-1-1/