

The Goals of Data Collection at Water Company United

Collect temporal **water usage patterns** among consumers → ***Predict demand at all times of the day***

Collect **water quality indicators** (pH, turbidity, chlorine, bacteria levels) at the water plant



Ensure compliance with laws for water quality

Install **acoustic, flow and pressure sensors** in pipes to catch leakages and blockages in pipes



Proactively catch technical problems and fix them efficiently



Collecting Water Usage Patterns

*Importance of data—
3.5/5
Difficulty of collection
and preparation for
analysis – 3/5*

- *Smart water meters, temperature sensors*
- Use the data to fit an **AutoRegressive model** for each household, which predicts the next water demand state using data from a few prior timepoints→
Supply optimization in the water plant through demand forecasting
- Accurate prediction of future consumption leads to **more efficiency** and **lowers prices** for the consumer

Collecting Water Quality Indicators

- *Installing pH and chlorine sensors is cheap and easy. Implementing other measures such as tracking bacteria requires collecting samples, but can be done on a regular basis.*
- Train a **logistic regression classifier** for real-time sensor data classification on expert annotated data to automatically assess if the water is safe for use.
- This data collection is critical due to business ethics and the necessity to comply to regulatory standards.

Importance of data—
5/5

Difficulty of collection
and preparation for
analysis – 2/5

Collecting Pipe Health Indicators

*Importance of data—
4/5
Difficulty of collection
and preparation for
analysis – 3/5*

- *Installing acoustic, flow and pressure sensors requires planning locations for sensor placement, cutting into the pipe and configuring the sensor to sync with the water plant.*
- Train an **unsupervised anomaly detection algorithm** for real-time sensor data on expert annotated data to automatically assess if the pipe needs fixing.
- Catching faults in pipes through sensors allows preventing outages and fixing technical problems while they are in their early phase, making it cheaper. Thus, this measurement modality is associated with savings for the company.

Project Prioritization and Summary

In summary, collecting different data modalities is associated with various levels of difficulty and the importance of the modality for the company also varies.

- Because sensors are digital it is easy to process on a computer. However, collecting other information such as bacteria contents of water can be associated with manual lab tests and is thus more difficult (however, this data is also crucial for **legal compliance**).
- The installation of sensors also varies in difficulty.

In consequence, taking account the difficulty of data collection and processing we rank the data in order of importance:

1. **water quality indicators**
2. **pipe health monitoring (flow, acoustic, pressure sensors)**
3. **tracking temporal water usage patterns**

Collecting this data is projected to yield savings and improved safety by using machine learning to automatically use the data for prediction and classification tasks.