

Команда «Семейная» - ЛЦТ 2025, задача №2. «Построение рекомендательного сервиса прогнозирования возникновения технологических ситуаций»

Текущая ситуация

Предиктивная аналитика по авариям, как мы считаем, на данном этапе не сильно нужна, т.к. в настоящий момент нет ни хорошо размеченных данных, ни статистики, ни корректных параметров. На предварительной стадии более оправдана только реактивная модель по инцидентам и предиктивная по потреблению ресурсов, но с максимально большим количеством параметров для поиска корреляций. На первом этапе нужна выстроенная обратная связь – система ищет потенциальные проблемы, а диспетчера обязательно проверяют указанные ИТП и вводят в систему результат (например, в виде текстовых отчетов для обработки большими языковыми моделями) для построения набора шаблонов действий. У заказчика с большой вероятностью может не быть опыта построения похожей системы и иметься недостаточно подготовленных данных, может не быть сильной мотивации из-за неочевидных преимуществ сервиса для стейкхолдеров в данный момент, нет выделенных ресурсов (людских и материальных), но, все-таки, решение нужно быстро и с максимальной экономической обоснованностью с точки зрения заказчика.

Как мы видим систему

Любая инженерная система, в том числе, Мосводоканал – это динамический граф, отражающий структуру системы, потоков воды, энергии, историю и текущее состояние оборудования всех видов, действий персонала и т.д. Поэтому решения с использованием теории графов являются логичным выбором, в том числе и в рамках машинного обучения и работы с большими данными.

Наше видение основных моментов при разработке «Рекомендательной системы...»:

1. Основа - граф системы холодного водоснабжения (Граф), начиная от станций водоподготовки (4 станции питьевой и 3 технической) и заканчивая МКД/промышленными и иными потребителями.
2. Узлы графов (Узлы) – точки измерения расходов воды/разветвления линий подачи ХВС с данными по замерам давлений/расходов в измерительном приборе.
3. Ребра графов (Ребра) должны содержать информацию о ключевых параметрах, влияющих на аварийность: длине, материале изготовления/покрытия, диаметре, рабочем давлении, даты начала эксплуатации и т.д. в соответствии с документом СХЕМЫ ВОДОСНАБЖЕНИЯ И ВОДООТВЕДЕНИЯ ГОРОДА МОСКВЫ ДО 2025

ГОДА и программой для ЭВМ: «Оценка надежности». (Входит отдельным модулем в программу «ЕИП-ЦТД»)), а также разность расходов ХВС на узлах графа, соединенных данным ребром.

4. Тупиковые узлы – измерительные приборы ХВС на входе в ИТП или при сдаче потребителю и измерительные приборы ГВС/ХВС на выходе из ИТП.
5. В качестве рабочих данных собираются все значения параметров для Узлов и Ребер Графа, а также данные по инцидентам/авариям на узлах/ребрах графа и принятых решениях и действиях по каждому инциденту со ссылкой на точку возникновения инцидента.
6. Для оптимизации размера и скорости работы математическая модель системы водоснабжения (Модель) имеет несколько модулей – **модуль мгновенного реагирования (ММР)** (работает только с мгновенными данными, получаемыми в реальном режиме, реагирует на признаки аварии, например, мгновенное падение или рост давления) и **модуль среднесрочного реагирования (МСР)** (работает с суммированными за определенный период (30 мин – 1 час) данными, реагирует на отклонения от нормальных паттернов/значений параметров) и **модуль выдачи рекомендаций (МВР)**.
7. Модуль ММР анализирует мгновенные расходы и давления (или иные параметры) в Узлах Графа и реагирует на заранее определенные уставки в отслеживаемых параметрах, которые считаются признаками инцидента/аварийной ситуации (падение давления до 0, резкий рост расхода выше рабочего порога и т.д.) – это простое решение, не требующее машинного обучения.
8. Модель МСР анализирует следующие основные параметры:
 - 1) разницу между расходом в материнском и суммами расходов в дочерних узлах графа в рамках одного дерева, исходя из предположения, что расход в любом узле более высокого ранга (А) должен быть равным сумме расходов в узлах более низкого ранга (В), непосредственно связанных с узлом высокого ранга (А);
 - 2) Разницу между расходами ГВС на ИТП и ХВС от ИТП;
 - 3) Паттерны потребления на каждом узле графа (за каждый час суток в зависимости от параметров: день недели, рабочий/выходной день, неделя года, сезонность, температура воздуха, тип потребителя или узла, район, станция водоподготовки и т.д.)
9. Модель (МСР) после обучения должна предсказывать расход на каждом узле на выбранный период вперед и выдавать уведомление в случае несоответствия реального паттерна потребления по п. 8.3. и исторически сложившегося/предсказанного моделью паттерна потребления. Пример такой модели и ее возможная архитектура предлагается в файлах решения.

10. Модуль (MCP) должен оценивать разницу (среднюю, накопленную и т.д.) в соответствии с п.п. 8.1 и 8.2. и выдавать уведомление в случае превышения разницы на заранее определенный уровень (5%, 10% и т.д.).

11. Считаем необходимым Модули кластеризации потребителей/ребер/узлов для выявления сходных по поведению объектов разных уровней и коррекции графовой модели для учета неочевидных, но выявляемых моделью связей для дальнейшей проработки и оценки причин таких связей.

12. Для визуализации и удобной работы с моделями разумным считаем надстройку в виде web-сервиса

13. Для выдачи рекомендаций диспетчерам мы предлагаем использовать RAG систему в связке с большими языковыми моделями для обработки сигналов от модулей MCP/MMP и выдачи рекомендаций/проектов решений на основе имеющихся стандартов/регламентов/отчетов и прочей документации Мосводоканала с указанием ссылок и приведение соответствующих цитат/выдержек из документации для облегчения принятия решения диспетчерами.

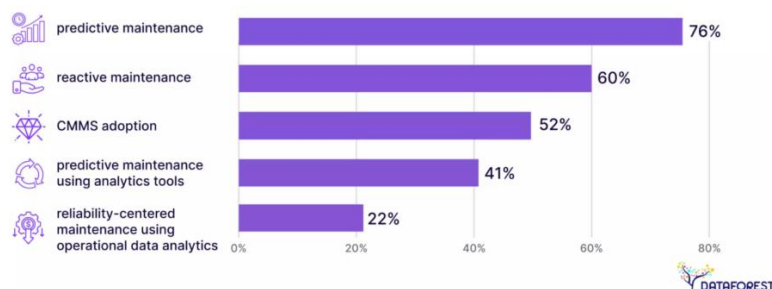
Цели и задачи

Экономический эффект

В соответствии с документом¹⁾ суммарный годовой ущерб от отказов трубопроводов Московского водопровода составил в 2011 году при средней стоимости ликвидации одной аварии (с учетом затрат на благоустройство) в 312, 57 тыс. рублей:

$$Y_{\Sigma} = 2524 \text{ отк./г.} \times 312570 \text{ р.} + 2104,74 \text{ млн.р} + 986,215 \text{ млн.р.} = 3873,625 \text{ млн.р.}$$

Статистические данные показывают, что при переходе от реактивной модели реагирования к предсказательной модели на каждом этапе происходит сокращение количества аварий/инцидентов на 15-25%, что в условия Москвы эквивалентно экономии порядка 700-900 млн. р. в год в ценах 2011 года.



Дополнительно, предиктивное обслуживание увеличивает ожидаемый срок службы оборудования на 20-40% (Predictive Maintenance for Utility Services – Preventing Disruptions).

Прогнозирование паттернов расходов холодной воды может помочь повышению энергоэффективности Мосводоканала за счет использования частотных приводов насосного оборудования и превентивного снижения мощности насосов и потерь энергии на прокачку в периоды сниженного водопотребления в сегменте системы

Социальный эффект

Интегрированный сервис, отслеживающий обращения граждан, дополнительно улучшает восприятие действий Правительства и ЖКХ Москвы жителями города за счет быстрого реагирования и решения проблем с водоснабжением и повышения уровня жизни в городе.

Этапы построения сервиса

Этапы работ (фазы, с ожидаемыми результатами)

Фаза А — Исследование и проектирование

- Собрать требования от заинтересованных лиц: диспетчерская, эксплуатация, аналитики, ИТ-безопасность.
- Определить источники данных и схему передач (форматы, протоколы, доступность метаданных).
- Определить «ground truth» — как мы будем получать метки порывов (журналы аварий, заявки, опросы бригад).
- Результат: спецификация данных, приоритетные метрики, минимальный пилотный набор ИТП.

Фаза В — Пилотный сбор и инфраструктура данных

- Развёртывание сегмента Мосводоканала для пилота (например все ИТП одной станции водоподготовки или иное).
- Компоненты: шлюз → брокер сообщений → стриминг-процессинг → долговременное хранение данных.
- Настроить нормализацию, валидацию, базовую очистку (пропуски, выбросы, дубли).
- Результат: работающий конвейер для пилотного проекта.

Фаза С — Система аномалий расходов воды в реальном времени

- Реализовать набор базовых детекторов аномалий (пороговые, статистические, сезонные) для оперативных оповещений.
- Интеграция с панелью мониторинга и каналами оповещений (SMS, e-mail, тикетинг).
- Результат: уменьшение времени отклика на аномалии.

Фаза D — Машинное обучение и предсказание порывов

- Собрать исторические метки (порывы) и объединить с временными рядами + внешними данными.

- Построить базовые модели (логистическая регрессия / XGBoost) + продвинутое (сверточные сети, LSTM/Transformer, графовые нейронные сети, для этого необходима топология сети).
- Внедрить pipeline обучения, валидации и мониторинга моделей.
- Результат: модель вероятности порыва с метриками (AUC, precision и т.д.).

Фаза Е — Производство, масштабирование, UX

- Масштабирование до всех 100k ИТП, кластеры хранения, кэширование агрегатов.
- Финальные дашборды/отчёты (табличные выгрузки, интерактивные графики, фильтры).
- Роли, права доступа, SLA на ответы API/отчёты.
- Результат: полнофункциональная система с поддержкой 5000 пользователей.

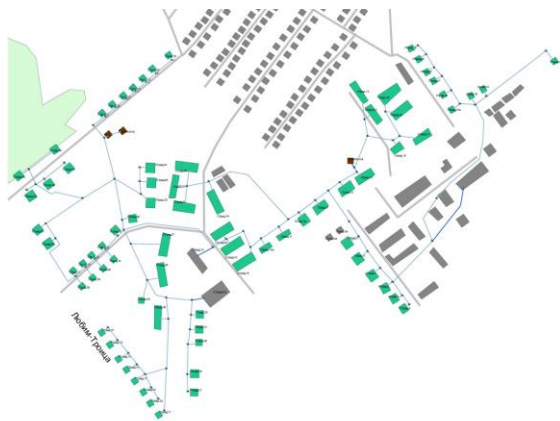
Дополнительные полезные источники данных (существенно повышают качество прогноза)

1. **Метеорологические данные:** температура воздуха, осадки, циклы замораживания/оттепели — влияют на морозные повреждения.
2. **Почвенные/геотехнические данные:** уровень грунтовых вод, подвижки грунта, промерзание почвы.
3. **Топология и атрибуты сети (GIS):** материал труб, возраст, диаметр, глубина прокладки, дата последнего ремонта, конфигурация сети (для GNN).
4. **Давление в сети:** если доступны датчики давления — очень ценная информация.
5. **Трафик/строительство:** дорожные работы, земляные работы рядом (пермиты/информация от муниципалитета).
6. **История гидравлических операций:** переключения, гидравлические испытания, чистки.
7. **Сигналы от систем SCADA/PLC:** насосы, клапаны, операции ИТП.
8. **Жалобы абонентов / кол-центр** (гео + время).
9. **Ремонтные бригады / план работ** (плановые отключения и раскопки).
10. **Качество воды / хим. параметры** — резкие изменения могут указывать на внешние воздействия.
11. **Спутниковые/аэросъёмки:** в некоторых кейсах термография и датчики могут помочь (но это опционально и дорого).

Дополнительные комментарии и подходы к анализу данных/построению модели

Система холодного водоснабжения любого населенного пункта представляет собой сложный граф, чаще всего, направленный, в виде дерева, хотя может иметь и другую структуру. Каждое ребро графа может представлять собой точку разветвления линий водоподачи или точку замера (место установки измерительных приборов), имеющую определенный вектор параметров (расход, местоположение и т.д.). Каждый узел — это участок трубопровода или некое техническое устройство/объект, через которые происходит перемещение потока воды и который

также может быть представлен в виде вектора (давление, материал изготовления, длина, глубина заложения, диаметр, свойство грунтов, защита и т.д.). В идеальном случае основной параметр для всех узлов/ребер – расход воды – должен быть однозначно вычислим в любой момент времени. Поскольку вода несжимаема и относительно изотермична, а также мы считаем, что потерь/утечек нет, то расход входящей в граф воды должен быть равен сумме расходов для каждого набора узлов одинакового ранга (расстояния от входного узла). Расходы в системе меняются во времени, в зависимости от времени суток, дня недели, сезона и т.д. Для такой системы в настоящее время наиболее эффективным методом предиктивной аналитики могут быть темпорально-пространственные графовые сверточные нейронные сети



При составлении графа – следующие признаки для линий водоснабжения в признаки узлов/ребер:

На основе результатов работы программы для ЭВМ «Оценка надежности» для оценки показателей надежности трубопроводов Московского водопровода к косвенным внешним факторам влияния на уровень надежности трубопровода и следовательно риску возникновения его отказа были выделены:

- материал труб,
- наличие и качество изоляционного покрытия,
- возраст трубопровода (год укладки трубопровода),
- защита от электрокоррозии,
- диаметр трубопровода,
- коррозионная активность грунта,
- гидравлические характеристики,
- тип (характер) грунтов,
- величина давления (напоров) в сети,

- число зафиксированных повреждений и аварий на участке трубопровода,
- интенсивность транспортных потоков,
- наличие и глубина залегания подземных вод,
- глубина заложения труб,
- роль участка в системе водоснабжения.

Модель Мосводоканала

Предлагаем строить предиктивную модель системы водоснабжения на основе GNN/STGCN (Графовые сети/темпорально-пространственные графовые сети).

Модель предсказания потребления ресурсов (паттерны потребления) – сверточные нейронные сети + трансформеры + линейная регрессия. Дополнительно – XGBoost/решающие деревья для сравнения.

Кластеризация узлов и ИТП/Потребителей – стандартные методы (например, K-neighbors, DBSCAN etc.) для эмбедингов, представляющих ИТП (получение эмбедингов узлов/потребителей/ИТП при помощи энкодеров на основе сверточных сетей/Трансформеров)

RAG/LLM/Streamlit/Telegram – стандартная связка для организации чат-ботов, но наилучшая конфигурация будет обсуждаться после ознакомления с имеющимися документацией и информационными системами Мосводоканала

Мобильное приложение для повышения эффективности работы полевых специалистов и диспетчеров (доступ к отчетам/подготовка отчетов/геолокация/распознавание и т.д.) – также очевидный вариант развития сервиса.