Компетенции департамента перспективных разработок

Построение и внедрение аналитических систем (BI, + BigData)

1. **Проектирование и построение классического хранилища структурированных данных** (DWH, OLAP). SAP BI, Postgres
2. **Проектирование и построение распределенных хранилищ слабоструктурированных данных** (DataLake). экосистема Apache Hadoop
3. **Организация информационных потоков** (ETL). SAP BI/BO/XI, Apache NiFi, AirFlow
4. **Визуализация данных и построение аналитической отчетности.**SAP BI/BO, Qlik, OpenSource, собственные решения

Опыт:

1. **ИУС П ГИРГ**: построение системы аналитической отчетности для МРГ
2. **АСКО**: автоматизация отчетных форм для ПАО Газпром
3. **СППУР**: визуализация KPI для руководства МРГ
4. **ИУС ТОиР:** визуализация KPI процесса ТОиР

Data Science

1. **Анализ данных с применением статистического анализа и математического моделирования** (EDA). Анализ временных рядов (SARIMA), кластеризация (SGD)…
2. **Построение прогнозных моделей с применением методов машинного обучения** (ML). Классификация, регрессия, кластеризация
3. **Решение задач оптимизации с применением математического аппарата** (MinMax). VAR, FERMA, SAP APO
4. **Разработка рекомендательных систем на базе методов** (DSS).
5. **Глубокое обучение с применением нейронных сетей** (DL) Опыт: NLP:NER,s2s, CV

Опыт:

1. **Прогноз (для АСПДУ)**:
   - кластеризация конечных потребителей газа по их профилю потребления
   - прогнозные модели потребления газа
   - моделирование утечек газа
   - Речевая модель (текст-в-голос) для автооповещения
2. **ИУС ТОиР:**
   - Прогнозная модель аварийности оборудования
   - AR-технологических схем на реальный объект
   - идентификация сущностей (адреса, запчасти, аналоги) в технической документации ТОиР
   - идентификация дефектов по фото-видео данным
3. **Автосекретарь** NLP инструмент для мониторинга инфо-пространства
4. **Риски:** модель оптимизации распределения операционных рисков

Разработка дополнительной функциональности SAP ERP (ABAP)

1. **Разработка ABAP.** Опыт: для модулей FI, FI-AA, SD, MM, CO, PM, PS, FI-FM, XI, RE, PP, PM, HCM (PA, PT, PY, LSO), RE
2. **Внедрение функционала SAP RCM.** Опыт: автоматизация процессов управление договорами, юридически значимый документооборот
3. **Настройка и разработка информационных потоков на базе сервисной шины SAP.** Опыт: SAP PI/XI (экстракторы, веб-сервисы и т.д.)

Разработка Веб и Мобильных приложений

1. **Разработка web-сервисов с применением микросервисной архитектуры.** Опыт: SAP, OpenSource (создание REST API, SOAP, oData сервисов)
2. **Создание «тонких»(браузерных) клиентов для «энтерпрайз» решений.** Опыт: SAP UI, ReactJS
3. **Разработка мобильных приложений для iOs и Android устройств.** Опыт: React-Native

Опыт:

1. **мТОиР**: мобильный клиент для ремонтных бригад
2. **мАДС**: мобильный клиент для аварийных служб
3. **МПБ**: веб-приложение для мониторинга передвижения бригад
4. **«Тонкие» клиенты для различных микросервисов**

Организация хранилищ структурированных данных
подход Кимбалла

Применяем при необходимости соединения разрозненных денормализованных витрин в единое хранилище данных, таким образом обеспечиваем доступ к актуальным данным с минимальными задержками, объемами хранимых данных (в случае виртуального хранилища данных).

ОГРАНИЧЕНИЯ: Существенная нагрузка на системы-источники данных, необходимость их доступности и ограничения на объем используемых данных.

КОНЦЕПЦИЯ BOTTOM-UP ИЛИ ПОДХОД КИМБАЛЛА (RALPH KIMBALL)

ПРОДУКТЫ

ПРОЕКТЫ

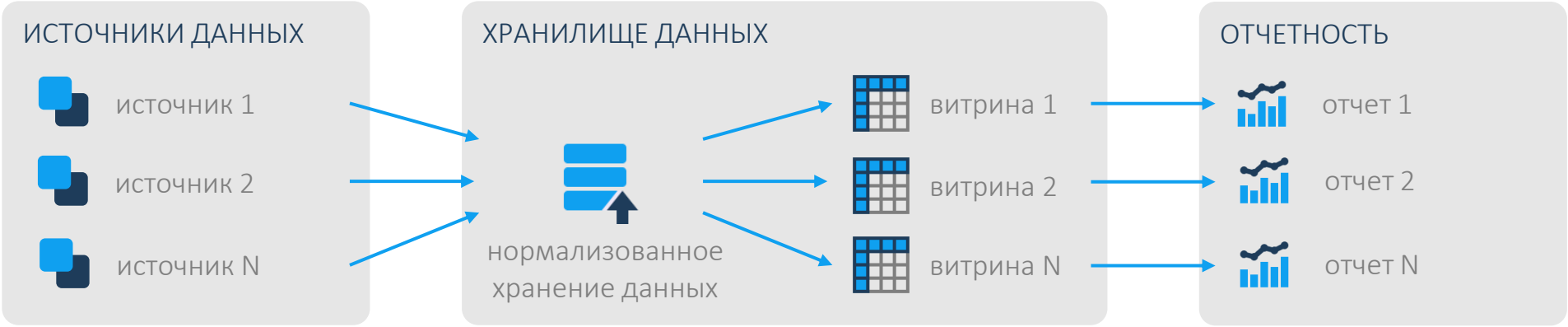
- Аналитика и отчетность ИУС ТОиР (МПБ)
- Дашборды

Организация хранилищ структурированных данных
подход Инмона

Применяем при необходимости построения единого централизованного места хранения данных и порождении пользовательских витрин данных только из него, что позволяет построить объектную модель предметной области бизнеса и проекцию этой модели на информационный уровень.

ОГРАНИЧЕНИЯ: Требует дополнительных затрат на процессы трансформации и загрузки данных, а так же ресурсы на их хранение.

КОНЦЕПЦИЯ TOP-DOWN ИЛИ ПОДХОД ИНМОНА (BILL INMON)



ПРОДУКТЫ

ПРОЕКТЫ

- BI система КИС МРГ
- BI система ИУС П ГиРГ

Организация хранилищ слабоструктурированных данных с применением концепции DataLake

В случае необходимости организации централизованного хранения данных, но с большим акцентом на разнородность, слабую структурированность и изменчивость данных применяем концепции Data Lake и Data Vault

ЛЯМБДА АРХИТЕКТУРА

ПАКЕТНЫЙ УРОВЕНЬ
архив сырых данных

СЕРВИСНЫЙ УРОВЕНЬ
обработка данных

ПОТОКОВЫЙ УРОВЕНЬ
данные реального времени

КОНЦЕПЦИЯ DATA VAULT

ИСТОЧНИКИ ДАННЫХ

- источник 1
- источник 2
- источник N

ХРАНИЛИЩЕ ДАННЫХ

хранение сырых детальных данных

нормализованное хранение данных

- витрина 1
- витрина 2
- витрина N

ОТЧЕТНОСТЬ

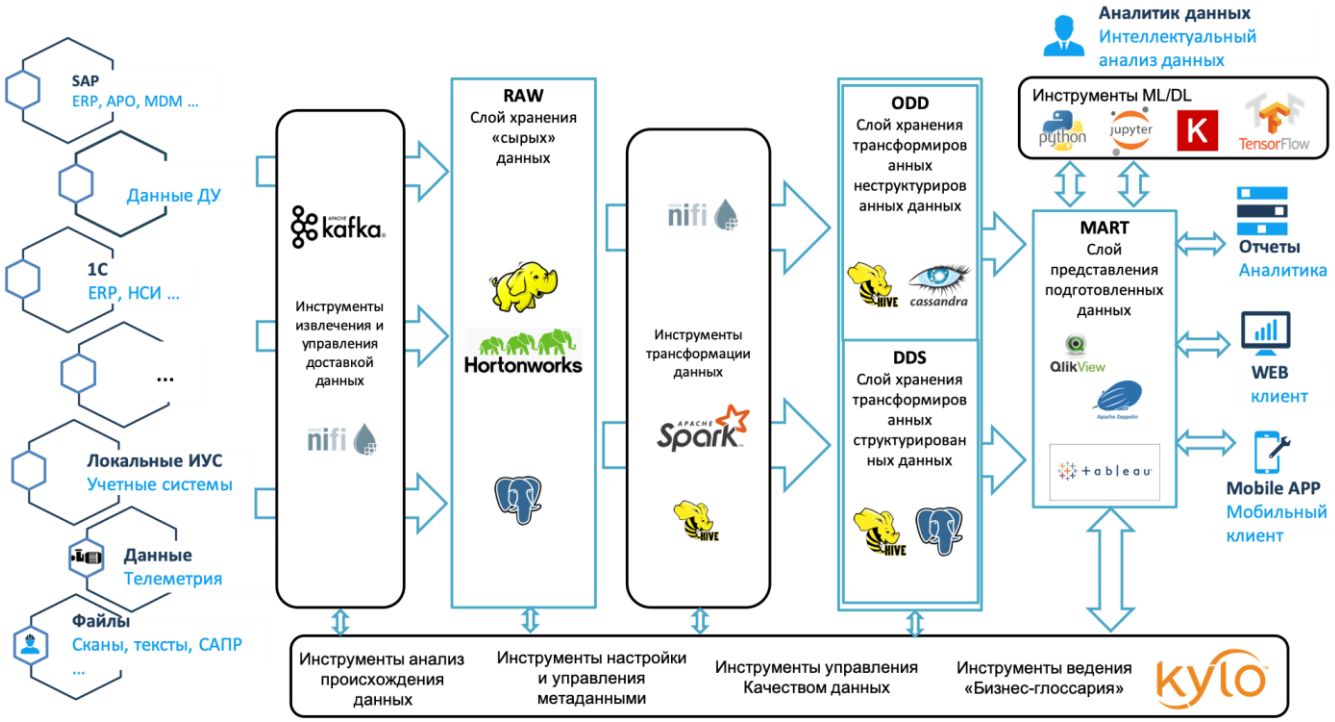
- отчет 1
- отчет 2
- отчет N

ETL инструмент	Брокер сообщений	Распределенное Хранение данных	Распределенная обработка данных	Инструменты машинного обучение	Доступ к данным в HDFS в виде SQL	Распределенная NoSQL база данных	Ad-hoc анализ и визуализация

Анализ Больших Данных

АРХИТЕКТУРА

для проведения анализа данных и внедрения инструментов машинного обучения в бизнес-процессы заказчика мы используем подходы по построению Лямбда-архитектуры и организации хранилищ по принципам Data Vault



МЕТОДОЛОГИЯ

для организации процесса построения математических моделей и интеллектуального анализа данных нами применяется методология CRISP-DM

ПРОЕКТЫ

- Предиктивный анализ объемов поставки газа в разрезе точек подключения;
- Предиктивный анализ выхода оборудования из строя или перехода из одного штатного режима в другой;

Визуализация данных и построение аналитической отчетности

НАША ЭКСПЕРТИЗА

включает использование инструментов от лидеров BI систем:

SAP Qlik

tableau

А так же собственные решения на базе open-source:

Apache Zeppelin

Superset

d3.js

Dash

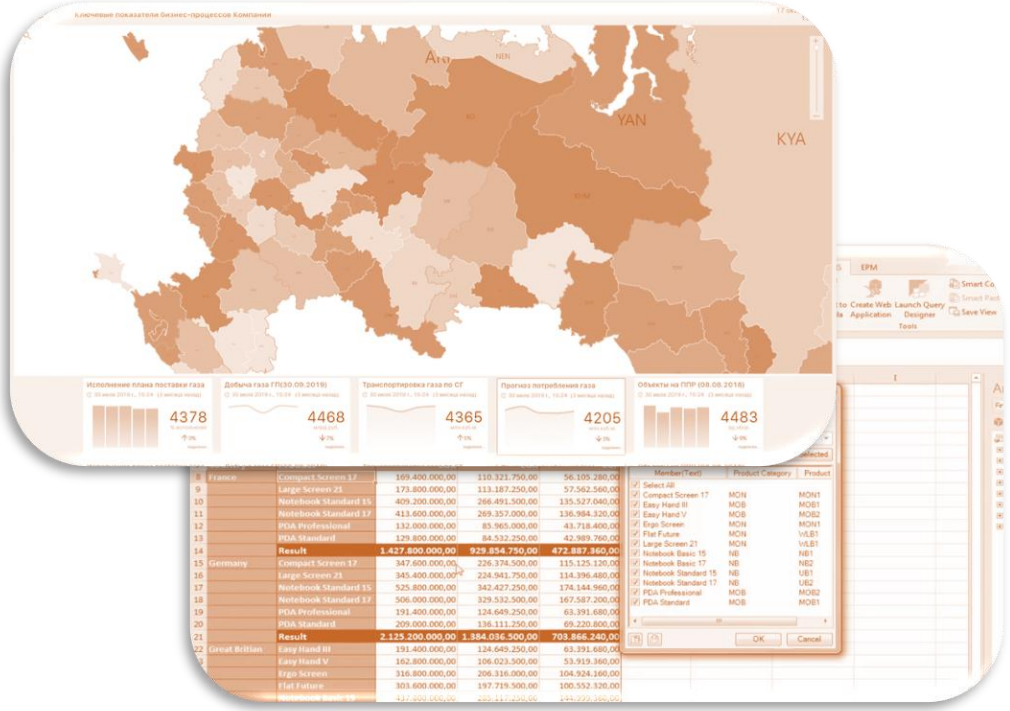
plotly

Altair

Регламентная отчетность

Аналитическая отчетность

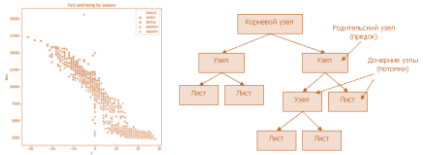
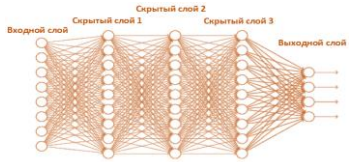
Информационные панели



Data Science

Математическое моделирование с использованием методов машинного обучения

для построения моделей с использованием методов машинного мы используем следующие подходы и алгоритмы:

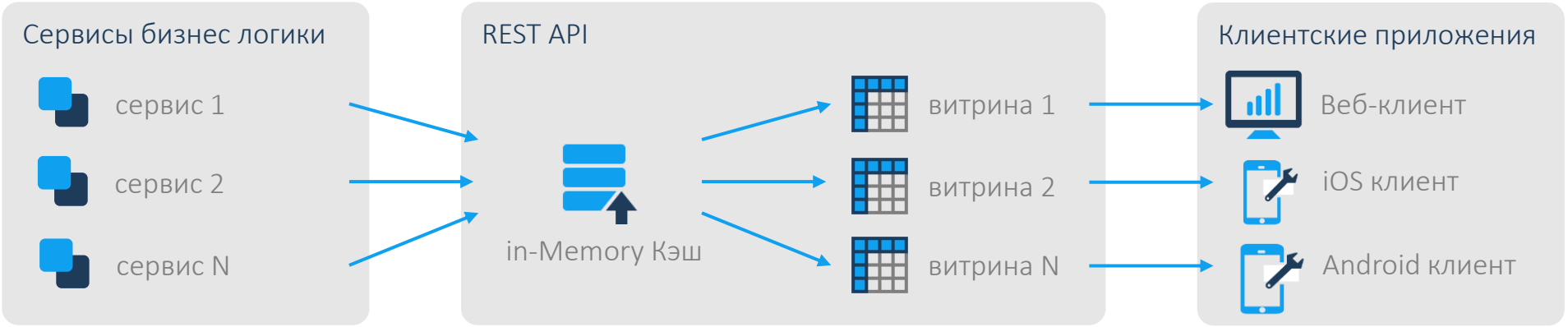
МОДЕЛИ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ	КЛАССИЧЕСКИЕ АЛГОРИТМЫ	ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ	НЕЙРОННЫЕ СЕТИ
			
Предиктивная аналитика, идентификация аномалий, выявление трендов, сезонности в данных	Предиктивная и прескриптивная аналитика, классификация, восстановление зависимостей	Кластеризация, идентификация аномалий, тематическое моделирование	Анализ фото-видео контента, анализ текстовой информации, работа со звуком, предиктивная аналитика
алгоритмы			
▪ ARIMA, SARIMAX ▪ GARCH	▪ Линейная регрессия ▪ Логистическая регрессия ▪ Бустинг и бэггинг над деревьями решений	▪ K-means ▪ FOREL ▪ Метод главных компонент (PCA) ▪ t-SNE	▪ RNN (LSTM) ▪ CNN
+			
▪ высокая степень научной обоснованности прогноза ▪ высокая точность интервальных прогнозов	▪ простота и прозрачность моделирования ▪ скорость работы моделей ▪ способность работы с большим набором признаков ▪ отбор наиболее важных для прогноза признаков	▪ работа с неразмеченным набором данных ▪ понижение размерности данных ▪ отбор наиболее важных для прогноза признаков ▪ высокая интерпретируемость тематического моделирования	▪ идентификация неочевидных зависимостей ▪ нелинейность и адаптивность
−			
▪ высокие требования к исходным данным ▪ необходимость моделирования каждого объекта ▪ очень низкая адаптивность модели	▪ необходимость формирования большого количества признаков для учета разных факторов ▪ высокие требования к СТИ, особенно к оперативной памяти	▪ необходимость априорных знаний о характеристиках кластеров ▪ неустойчивость алгоритмов ▪ высокие требования к СТИ	▪ высокие требования к исходным данным ▪ невозможность интерпретации алгоритма формирования прогноза («черный ящик»)

Разработка Веб и Мобильных приложений

Функционал разработанных приложений мы реализуем путем создания атомарных сервисных приложений, которые «безболезненно» встраиваются в любую IT-инфраструктуру Заказчика. Для разработки клиентской части мы используем технологический стек React/React-Native, который позволяет создавать платформо-независимые приложения как для веб, так и мобильных устройств.

ОГРАНИЧЕНИЯ: Накладывает ряд ограничений на использование специализированных функциональных возможностей устройств.

КОНЦЕПЦИЯ МИКРОСЕРВИСНОЙ АРХИТЕКТУРЫ



ПРОДУКТЫ

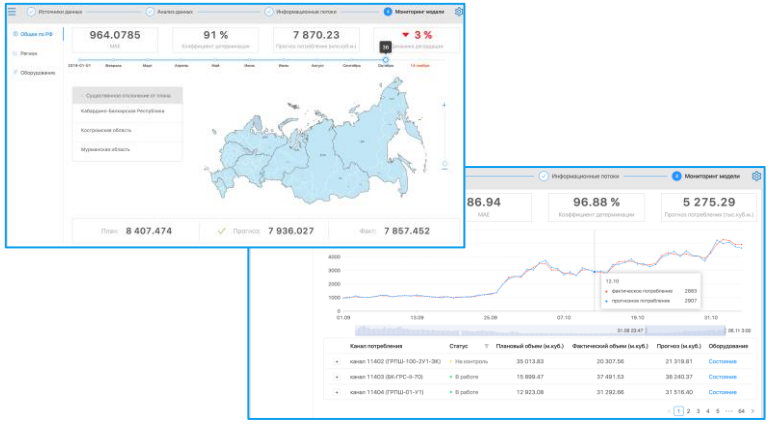
ПРОЕКТЫ

- мТОиР
- мАДС
- МПБ

Платформа Обработки Больших Данных **ГЕКСАГОН** является фундаментом для построения систем хранения и анализа данных. Обладает следующей модульной структурой:

ВСПЛЕСК	Модуль выявления аномалий и закономерностей в данных. На базе модуля может быть создана система проверки качества как справочных данных, так и любых информационных потоков Организации
ПРОГНОЗ	Модуль предиктивного анализа. На базе модуля может быть создана Система прогнозирования объемов поставки товара, Система прогнозирования выхода оборудования из строя и т.д.
СОВЕТНИК	Модуль рекомендательных систем. На базе модуля может быть создана Система поддержки принятия решений
ВЕСТНИК	Модуль голосового и текстового оповещения (в т.ч. чат-боты). На базе модуля может быть создана Система аврийно-диспетчерского оповещения, Система взаимодействия с клиентом с интеллектуальным подбором ответов и рекомендаций.
ЧТЕЦ	Модуль анализа текстовой информации(NLP). На базе модуля может быть создана Система анализа инфомедиа пространства, автоматический анализ технической документации и т.д.
ТЕЛЕСКОП	Модуль обработки фото-видео изображений(CV). На базе модуля может быть создана Система идентификации объектов на фото и видео

Реализация прогнозной модели поставки газа конечным потребителям с применением методов машинного обучения



ЦЕЛЬ

- Повышение эффективности газораспределения за счет увеличения точности планирования поставок и сведения баланса газа в разрезе конечных потребителей

ЗАДАЧИ

- Долгосрочный прогноз поставки газа в разрезе регионов и в целом по РФ (месяц, квартал)
- Прогноз потребления газа конечными потребителями (3, 5, 10 суток)

ФУНКЦИОНАЛЬНАЯ СХЕМА

РЕАЛИЗАЦИЯ

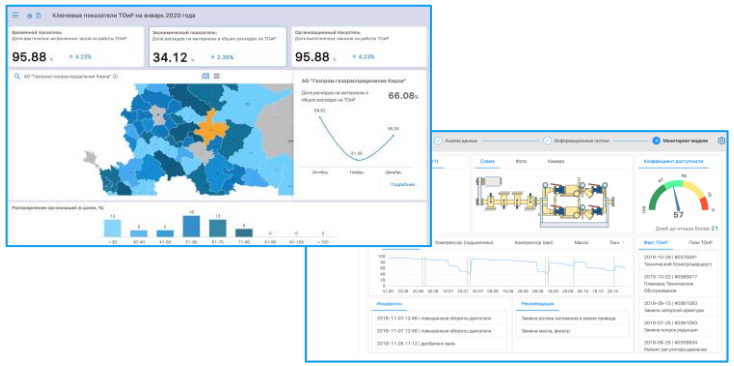
- ВХОДНЫЕ ДАННЫЕ:
  - Данные приборов учета газа (телеметрия) как исторические, так и в режиме реального времени
  - Прогноз погоды
  - Информация о потребителе (организационная форма, цель потребления, профиль потребления, договоры, платежи, …)
- МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ИСПОЛЬЗОВАННЫЕ ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ:

  K-means, PCA, SARIMAX, SVM, Градиентный бустинг над деревьями решений

по метрикам MAPE, accuracy на фактических данных

Точность прогноза **87-93** %

Разработка прогнозной модели доступности оборудования с применением методов машинного обучения



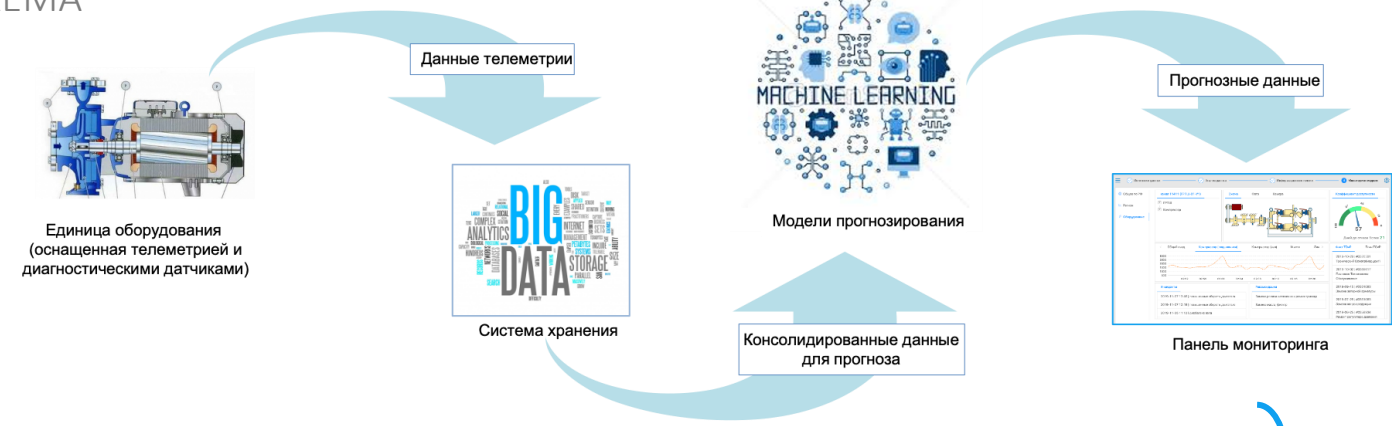
ЦЕЛЬ

- Повышение эффективности ТОиР за счет выбора оптимальной стратегии
- Снижение затрат на ТОиР за счет перехода на обслуживание «по состоянию» (с учетом надежности)

ЗАДАЧИ

- Расчет индекса доступности технического оборудования (OEE-анализ)
- Расчет и идентификация потенциального отказа оборудования (PF-интервал)
- Прогноз возникновения аварийных ситуаций на горизонте достаточном для принятия управленческого решения

ФУНКЦИОНАЛЬНАЯ СХЕМА



РЕАЛИЗАЦИЯ

- ВХОДНЫЕ ДАННЫЕ:
  - Исторические данные по аварийности и ТОиР
  - Диагностические данные (телеметрия, результаты диагностики, осмотров)
  - Информация о единице оборудования (режимы работы, роль в тех.процессе, техкарты,…)
- МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ИСПОЛЬЗОВАННЫЕ ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ:

  K-means, PCA, SARIMAX, SVM, Градиентный бустинг над деревьями решений, LSTM

по метрикам MAPE, accuracy на фактических данных

Точность прогноза **79-85** %

Подсистема визуализации показателей «Дашборды»

Дашборд — Информационная панель, позволяющая оперативно отслеживать состояние бизнеса посредством наглядной демонстрации ключевых показателей

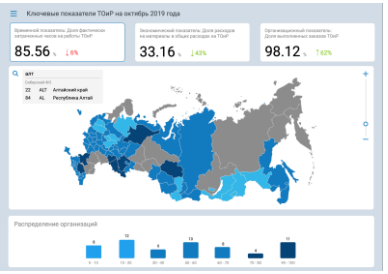
Для кого: — руководители среднего и высшего звена

Цель внедрения: — оперативное получение достоверной информации, выявление неочевидных взаимосвязей, контроль процессов для эффективного принятия бизнес-решений

Включает:

Визуализацию

- ✓ Лаконичное отображение ключевых показателей с использованием инфографики;
- ✓ Богатый набор фильтров, сортировок, средств поиска и других инструментов, позволяющих пользователям менять ракурсы отображения информации в широких пределах



Оперативное уведомление
о достижении пороговых значений

- ✓ Контроль достижения порогового значения ключевых показателей ;
- ✓ Заблаговременное предупреждение о развитии негативной тенденции;
- ✓ Отправка сообщений при помощи различных средств коммуникации (Email, SMS, мессенджеры, push-уведомления)

Отчетность и печать

- ✓ Печать инфопанели с графиками в формате «как есть» (как видно на экране);
- ✓ Формирование отчетов по установленной форме на основании данных, представленных в дашборде.

Конструктор дашбордов

Создание дашбордов собственными силами пользователей (без привлечения программистов):

- ✓ простая компоновка дашборда визуальными элементами с помощью функции Drag&Drop
- ✓ настройка пороговых значений для ключевых показателей дашборда, а также способов и средств уведомлений;
- ✓ привязка шаблонов для отчетных форм;