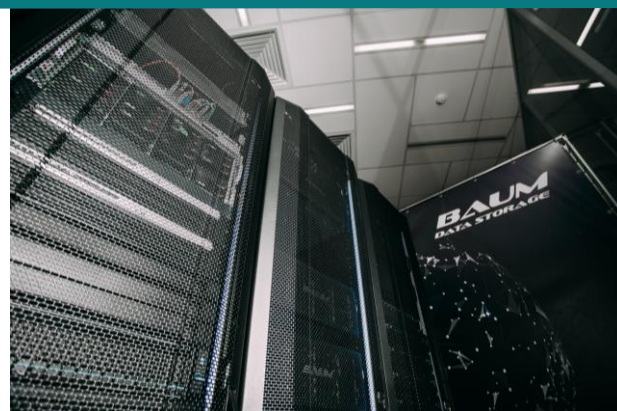


BAUM AI

ARTIFICIAL INTELLIGENCE

DATA STORAGE

BAUM AI



2010

Год основания

120+

Установок СХД и
платформы AI

250

Сервис центров



3

Центра R&D

2

Продукта –
Storage and AI

10+

Авторских курсов обучения
в МГТУ им. Н.Э. Баумана

BAUM UDS

Классическая унифицированная
система хранения данных

BAUM SWARM

Scale-out Storage

BAUM AI PLATFORM

Платформа прикладного
искусственного интеллекта

BAUM UDS



Возможности BAUM UDS

Deduplication Compression	Синхронная и Асинхронная Репликация	Масштабирование	
FC, iSCSI	Консистентный Snapshot VMapp	Версионность Квоты	Версионность Квоты
NFS SMB	ГОСТ Шифрование	WORM	Advanced Monitoring BAUM AI
Unified Storage	Snapshot Clone	Кэш до 60 TB NVDIMM, Optane, SSD	RDMA

Модельный ряд UDS



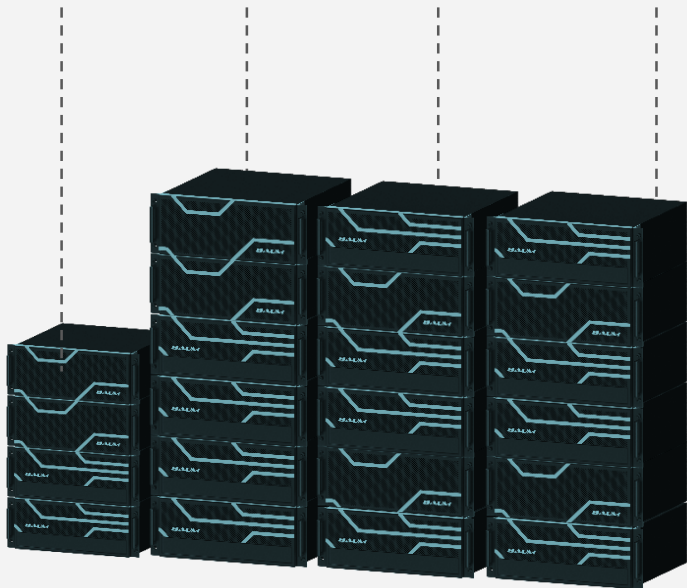
Mid-Range

UDS2000

UDS8000

UDS2500F

UDS4500F



Серв. СХД

БФ.М5.2/20



High-End NVMe

SPACE3500Nx



Сервис



B A U M

250

партнерских
сервисных центров

24/7

NBD

4

уровня
поддержки

An aerial photograph of a river system in a lush, green forested landscape. The river flows through a series of meanders, creating several oxbow lakes. The water is a light, milky color, and the surrounding land is covered in dense green trees. The sky is a pale blue with some light clouds. The image is framed by a large black and white geometric shape on the left side.

BAUM SWARM

Подход ВАУМ



1. Учесть опыт, накопленный в разработке линейки классических СХД корпоративного уровня.

1. Проанализировать сильные и слабые стороны современных СПО-решений.

3. Взять лучшие идеи из мира «классики» и СПО и создать решение «с нуля».

4. Адаптировать систему для российских аппаратных платформ и облачных платформ.



5. Ориентироваться на требования конкретных заказчиков сегодня и не иметь узких мест для развития завтра.

5. Интегрировать в дизайн системы все современные методики- CI/CD, сквозное модульное тестирование и т.п.

Ограничения существующих систем

Классические
модульные системы
хранения

проблемы с
пропускной
способностью
при большом
количестве
клиентов

оптимизированы
на классические
схемы

не
приспособлены к
горизонтальному
масштабировани
ю

географические
распределенные
сценарии либо
поддерживают только
часть функционала,
или вызывают резкое
удорожание
стоимости хранения

слабая
оптимизация
под HardWare

проблемы со
скоростью и
надежностью
работы

нет
подтвержденного
roadmap

требование наличия у
заказчика
собственной команды
разработчиков и
инженеров для
глубокой и
постоянной доработки
продукта

Open source решения

SWARM

Полностью оригинальная кодовая база ядра системы (отсутствует использование сторонних и opensource разработок)

Высокая производительность при параллельном доступе к данным

Быстрая линейная масштабируемость

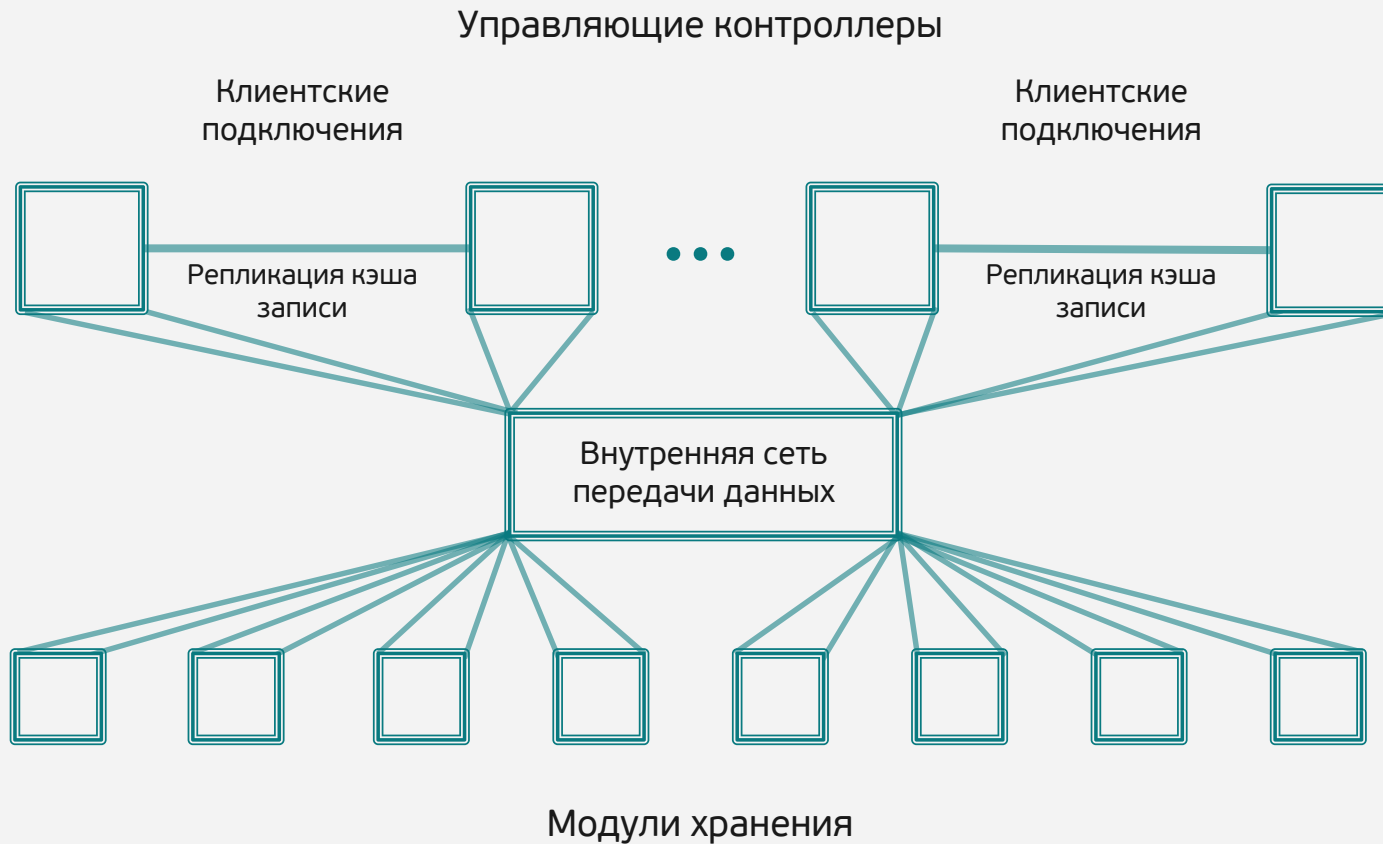
Протоколы:

NFS
SMB
S3
HDFS
REST
iSCSI
FTP
HTTP

Обеспечение SLA по качеству обслуживания для групп приложений

Дисковая отказоустойчивость и высокая доступность данных

Емкость хранения и пропускная практически не ограничена



Модуль управления «РАЗУМ»

взаимодействует с клиентами

«знает» и «отвечает» за себя,
связные модули управления и
множество модулей хранения

обеспечивает работу
внутренних служб кластера,
добавление
и замену модулей управления
и хранения

сервисы рассчитаны на
параллельную работу внутри
узла и эффективное
взаимодействие

управляет схемой защиты
данных и распределением
логических блоков данных по
узлам хранения

«на лету» **вычисляет**
размещение данных с учетом
резервирования и
восстанавливает целостность
при необходимости

работает как **кэш** записи и
чтения

пишет и читает данные с
модулей хранения, оптимизируя
крупные блоки данных под
последовательный доступ

работает с пулом своих
локальных дисков

«знает» только про себя и
кластер в целом

пишет и читает данные
во внутреннем представлении

проверяет целостность блоков
данных на своих дисках

восстанавливает ошибки
на своем уровне

сервисы ввода-вывода
ориентированы на
максимальный параллелизм в
рамках одного модуля

Модуль управления «МУСКУЛЫ»

A landscape photograph of a coastline with mountains and a lake, overlaid with a large white arrow pointing right. The text "BAUM AI PLATFORM" is centered within the white arrow.

BAUM AI PLATFORM

BAUM AI PLATFORM

Мы делаем программно-технический комплекс прикладного искусственного интеллекта BAUM AI, помогающий заказчикам из госсектора и среднего и крупного традиционного бизнеса быстро получать эффект от внедрения продвинутой аналитики на базе ИИ, используя сочетание оптимизированной аппаратной платформы с ориентированным на предметного пользователя, а также уникального софта собственной разработки, который позволяет создавать, обучать или использовать предобученные модели искусственного интеллекта без необходимости прямого кодирования по принципу drag&drop.

Алгоритм решения задач с помощью BAUM AI

Изучение задачи
заказчика
группой исследователей

Подготовка прототипа
при помощи платформы
BAUM AI

Помощь
по внедрению
в продуктив заказчика

Подготовка Roadmap
повышения точности
модели

Обучение специалистов
заказчика в МГТУ

Заказчик
получает решенную
задачу и возможность
использования кадров
из числа студентов и
научного персонала
МГТУ

Составные элементы и уникальные возможности

The screenshot displays the BAUM BPMN editor interface. The main workspace shows a BPMN diagram with the following elements: a Start event (Запуск), a Data Store (База данных), an Output Gateway (Шлюз выход), two Data Objects (Датасет), an Input Gateway (Шлюз вход), a Process (Процесс), and an End event (Конец). The process block is highlighted, and a configuration panel is open on the right.

time_series_shares.csv

Выберите: Столбцы для загрузки		
<input checked="" type="checkbox"/>	Без имени 1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	clock	<input checked="" type="checkbox"/>
	value	
0	11/01/10	226.9
1	12/01/10	226.7
2	13/01/10	225.6

Работа с данными и моделями по принципу drag&drop

«Сердце» платформы – универсальный конструктор AI с использованием блок-схем в нотации BPMN

ПТК AI – от загрузки датасета, выбора библиотеки и подстройки бизнес-процессов до графического результата за несколько кликов

Особенности Hardware BAUM AI



Адаптированный
кэш СХД

Оптимизированные настройки
доступа

RDMA

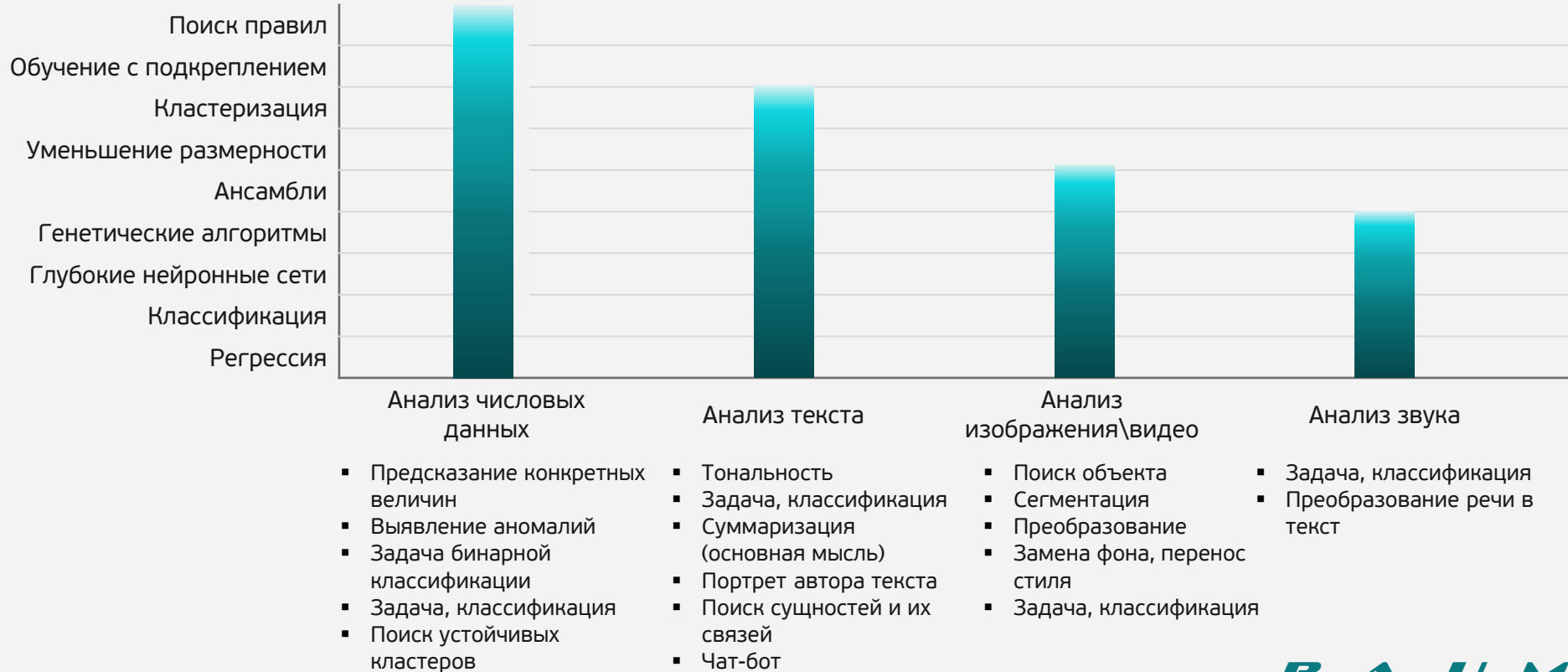
Спец. процессоры

Интеграция с Qos коммутаторов

Добавление контроллеров и
дисковых полок без остановки
системы и прекращения доступа
к данным

Оптимальный подбор алгоритмов и библиотек

Данные и алгоритмы



Перечень решенных кейсов



Типы данных



Временные ряды



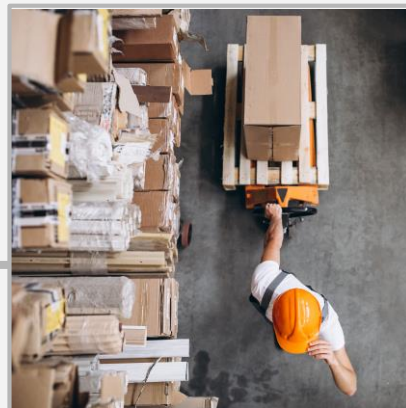
Мониторинг и
предиктивная аналитика
для ТЭЦ



Алгоритм определения
аномалий в банковских
транзакциях



Предиктивная аналитика
показателей серверов и
СХД



Определение ожидаемой
нагрузки на склад и
пиковых нагрузок



Предсказание лесных
пожаров

Текст



Цифровой профиль
должности



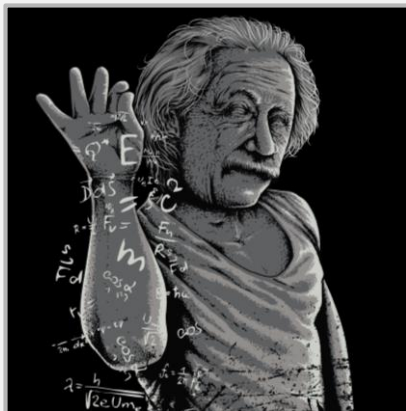
Определение автора по
тексту



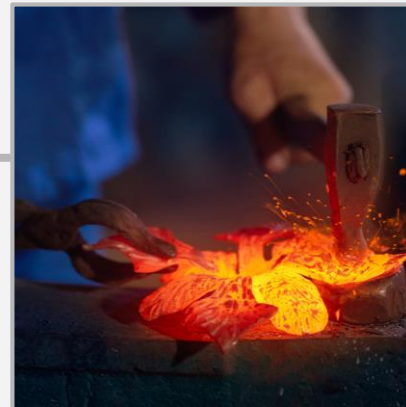
Интеллектуальная
обработка новостей



Разработка комплексной
СППР при анализе
входящей информации



Выявить тренды и
центры притяжения
научного внимания в той
или иной отрасли

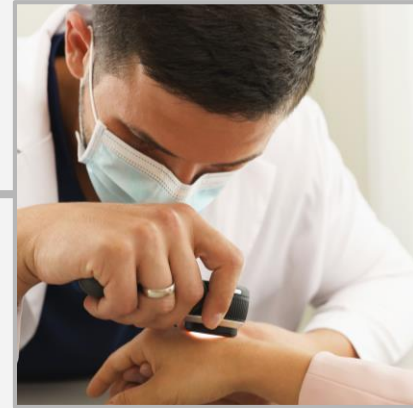


Создание
интеллектуальной базы
данных свойств
материалов и его
прогнозирование

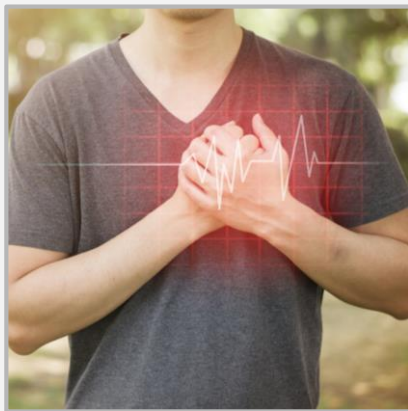
Изображения



Поиск изображений в
ГАРФ



Распознавание
злокачественных родинок



Классификатор сердечно-
сосудистых заболеваний



Распознавание дефектов
на металлических
изделиях



Фотоника – трекинг
частиц в суспензии

Звук



Определение объекта по
звуку



Классификация
музыкальных
аудиофайлов

Кейс 1

Временные ряды

Мониторинг и предиктивная аналитика для ТЭЦ

Задача

- Комплексный мониторинг состояния системы в реальном времени
- Уменьшение времени планового простоя
- Предиктивное предупреждение выхода из строя агрегатов
- Проведения только фактически необходимого технического обслуживания

Методы решения

- Объединение двух датасетов (звук и видео)
- Классификация (50 классов, 2000 образцов)
- Метод естественной синхронизации аудио и видео потоков (20 Тб)
- Подбор архитектуры нейронной сети SoundNet
- Обучение нейронной сети и тестирование

Результат



Алгоритм определения аномалий в банковских транзакциях

Кейс 2

Временные ряды

Цели

Создание алгоритма выявляющего признаки мошеннических транзакций.

Задачи

State-of-art алгоритмы машинного обучения:

- алгоритмы, которым в процессе обучения не показывали правильные ответы (аномалии);
- алгоритмы обучения с учителем, при которых аномалии заранее имели соответствующие метки.

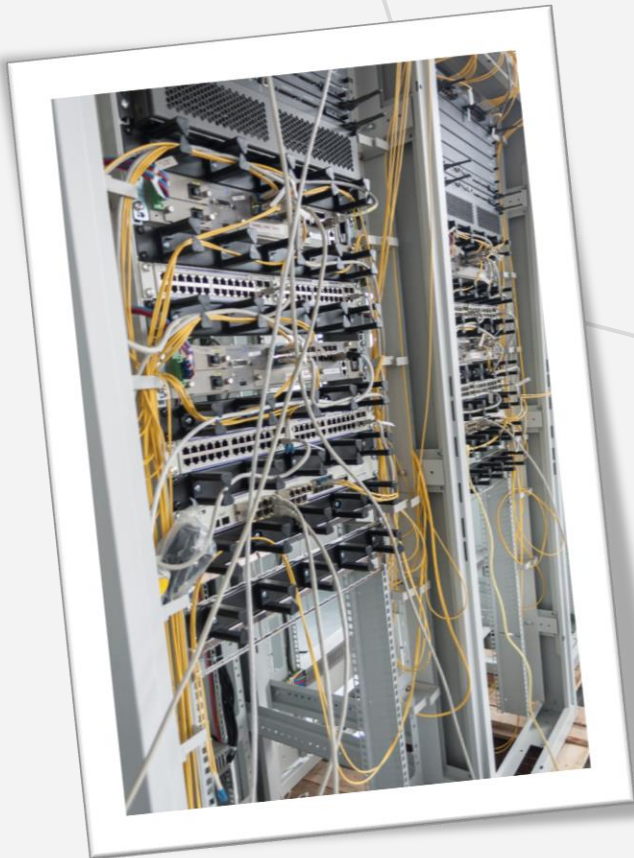
Результат

Возможность распознавать аномалии и максимально уменьшить вероятность блокировки нормальных банковских транзакций.



Кейс 3

Временные ряды



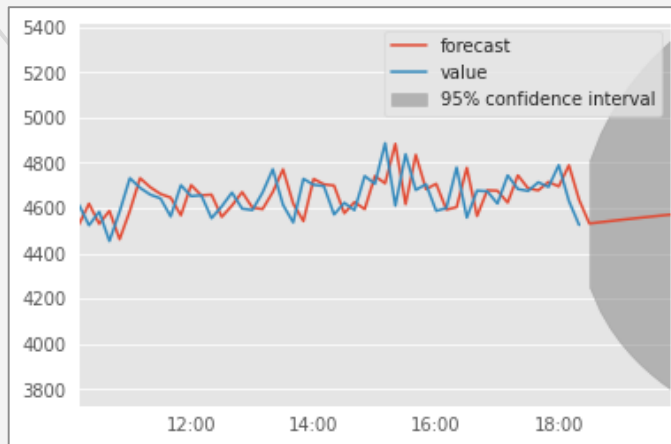
Предиктивная аналитика показателей серверов и СХД

Задача

- Предиктивно – статистический анализ временных рядов.
- Раннее предсказание показателей нагрузки на CPU, Latency, pool read / write iops.
- Zabbix – AI – Dashboard

Методы решения

- Нейронные сети: полносвязные слои, сверточные слои
- ARIMA - авторегрессия



Определение ожидаемой нагрузки на склад и пиковых нагрузок

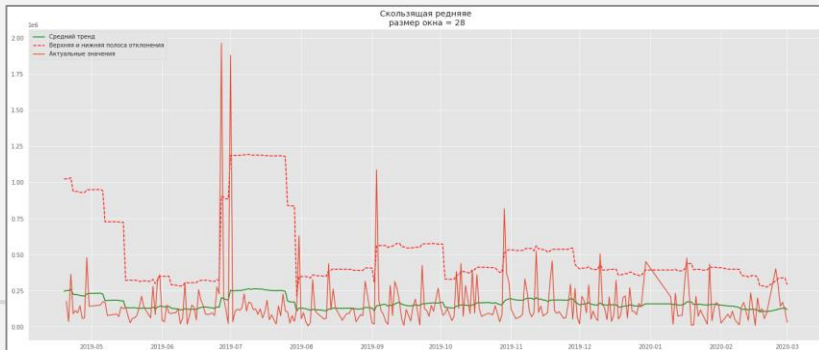
Кейс 4
Временные ряды

Задача

- По историческим данным, определить ожидаемое движение по складу (приход товара и отгрузка).
Спрогнозировать возможную пиковую нагрузку в течении месяца.

Методы решения

- Обработка временных рядов
- Нормализация и стандартизация
- Авторегрессия, стандартное отклонение, средние скользящие
- Нейронные сети с памятью (LSTM)



Кейс 5

Временные ряды



Предсказание лесных пожаров

Задача

Предсказание вероятности возникновения лесных пожаров, их координаты и площадь, используя исторические погодные данные, а также косвенные признаки (наличие инфраструктурных объектов, близость к транспортным сетям, социально-демографические характеристики ближайших населенных пунктов)

Методы решения

- Сбор данных и «склейка» в датасет
- Препроцессинг, извлечение геоданных
- Нейронные сети (Dense, LSTM, Conv1D)
- ML алгоритмы SVM, Autoencoders



Мониторинг параметров нефтепроводов. Охрана периметров и протяженных объектов

Кейс 4
Временные ряды

Цели

- Система предназначена для охраны и мониторинга объектов большой протяженности / площади

Задачи

- Обнаружить объект, Классифицировать эти объекты, выдавать сигналы тревоги в случае необходимости, отображать на карте,
- анализировать звуковые колебания с выбранных участков для дополнительного анализа на расстоянии от 5 до 100 метров

Результат

- Отображение в реальном времени активности вдоль кабеля в виде графика
- Возможность прослушивать звук с выбранных участков кабеля



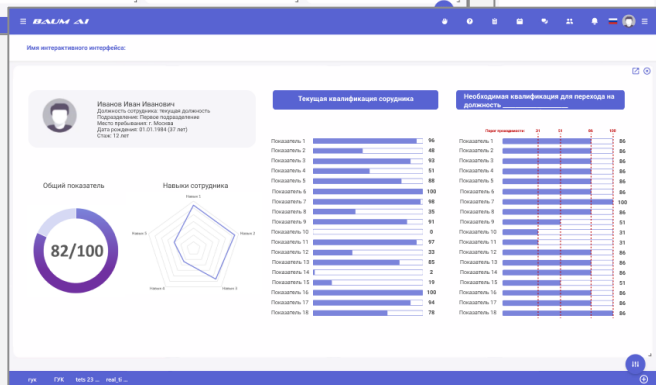
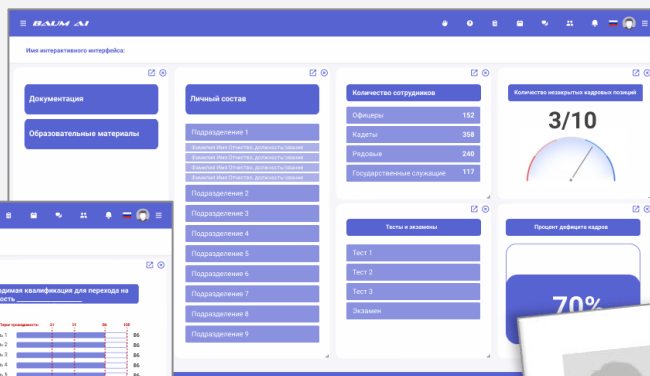
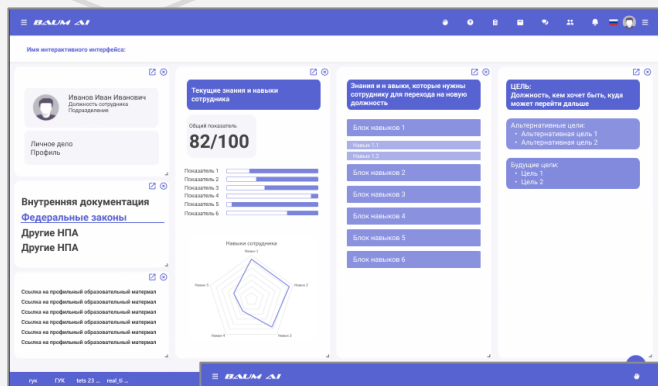
Кейс 1

Текст

Цифровой профиль должности

Задача

- определение персонального уровня знаний
- формирование индивидуальных образовательных траекторий
- оценка персонального потенциала в образовании



Определение автора по тексту

Кейс 2

Текст

Задача

Определение автора по тексту, с помощью рекуррентных сетей и одномерной свертки.

Методы решения

- Разработка функций обработки естественного языка (очистка от шума)
- Создание словаря (в т.ч. Bagofwords)
- Создание обучающей и проверочной выборки + embedding
- Подбор архитектуры нейронной сети (Dense, Conv1D, LSTM, maxpooling)
- Обучение нейронной сети, работа с переобучение

Результаты

- Достигнутые показатели:
- Точность распознавания текста 90,1%
- Перспективы применения:
- Применение данного и схожих методов возможны при разработке комплексных систем оценки соционики (психологический портрет авторов, их намерения и т.д.).



Кейс 3

Текст



Интеллектуальная обработка новостей

Задача

- Определение тональности (негативный, позитивный, нейтральный) новостного фона (10 000 новостей) за последние сутки
Извлечение сущностей (суммы, ФИО)
- Мониторинг национальных проектов
- Целевой пользователь – аналитик

Методы решения

- Разработка функций обработки естественного языка (очистка от шума)
- Создание словаря (в т.ч. Bag of words)
- Создание обучающей и проверочной выборки + embedding
- Подбор архитектуры нейронной сети (Dense, Conv1D, LSTM, maxpooling)
- Извлечение сущностей (Natasha)

Задача

Разработка классификаторов для принятия решений на основе потока открытых больших данных (текст, звук, изображения)

Методы решения

- Сбор и обработка больших массивов данных
- Разделение на классы
- Создание классов рекомендаций и их весов
- Подбор моделей анализа и классификаторов

Результаты

- Результаты научных исследований
- Нормативно-правовые документы
- Аналитические материалы
- Справочная информация
- Сообщения СМИ
- Числовые данные
- Сообщения от специализированных СМИ
- Извлечение эталонных данных (моделей) угроз
- Обработка и хранение данных оперативной информации
- Отчеты
- Интеллектуальный анализ данных
- Варианты предлагаемых решений



Кейс 5

Текст



Выявить тренды и центры притяжения научного внимания в той или иной отрасли

Задача

- разработать систему кластеризации и предиктивного анализа патентной базы
- выявить тренды и центры притяжения научного внимания в той или иной отрасли

Методы решения

- препроцессинг и структурирование массива реферативных данных
- извлечение из каждого реферата от 6 до 10 ключевых слов и их суммаризация
- определение основного смысла научной публикации
- группировка и кластеризация данных
- определение ключевых слов - областей максимального интереса

Создание интеллектуальной базы данных свойств материалов и его прогнозирование

Кейс 6

Текст

Задача

- создание интеллектуальной базы данных свойств материалов, полученных по технологии селективного лазерного плавления
- прогнозирование свойств полученного материала в зависимости от состава исходных компонентов и технологических режимов

Методы решения

- поиск, валидация, хранение, препроцессинг и анализ данных
- обучение и использование предобученных моделей ИИ

Результаты

- формирование базы данных физических и теплофизических свойств, механических и технологических характеристик полимерсодержащих материалов, как в условиях изготовления, так и в условиях искусственного старения
- многомерная модель базы данных, включающая описание свойств и характеристик материалов
- возможность выполнять первичную оценку свариваемости стали, оценивать количественные, технические и экономические показатели, значительно снижая потребность в создании опытных образцов



Кейс 1

Изображения

Поиск изображений в ГАРФ

Задача

- Распознавание и классификация лиц на старых фото
- Точность классификатора 85-90%

Методы решения

- Нейронные сети: полносвязные слои, сверточные слои
- Face recognition
- Bounding box



Сервис

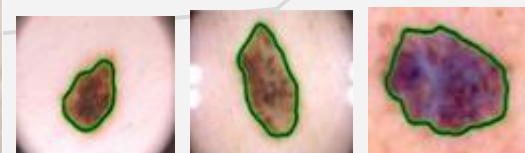
- Нейронная сеть для автоматической диагностики родинок на злокачественность. Пациент имеет возможность отправить фото своей родинки и получить оценку онлайн, далее оформить запись на приём.

Типы:

- melanoma (mel)
- melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc)
- Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)
- basal cell carcinoma (bcc)
- benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl)



Segmentation



Точность модели нейронной сети до 93,21%

Кейс 3

Изображения



Классификатор сердечно-сосудистых заболеваний

Сервис

- Классификатор сердечно-сосудистых заболеваний.
- Постановка диагноза в момент лабораторных испытаний.
- Онлайн уведомление пациента и врача.

Возможности

- Сбор данных по анализам с учетом дополнительных демографических
- Нейронная сеть + машинное обучение для постановки диагноза
- Отслеживание показателей анализов во времени и формирование рекомендаций пациенту

Распознавание дефектов на металлических изделиях

Кейс 4
Изображения

Задача

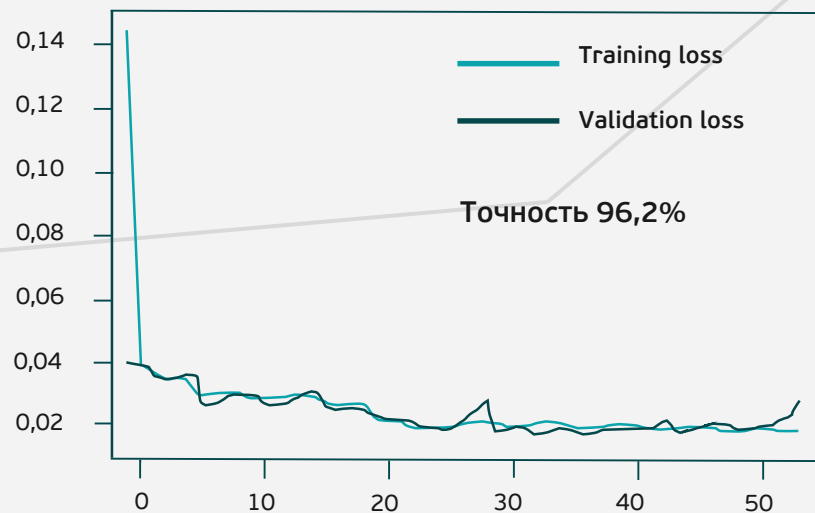
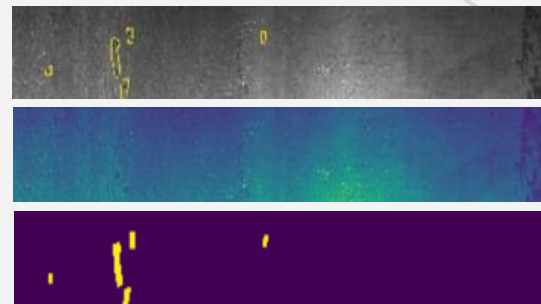
- Раннее распознавание микро дефектов на металлических поверхностях, с помощью нейронных сетей.
- Точность классификатора $> 0,9$

Методы решения

- Нейронные сети: полносвязные слои, сверточные слои
- Борьба с переобучением
- Дообучение нейронной сети
- Бинаризация изображений

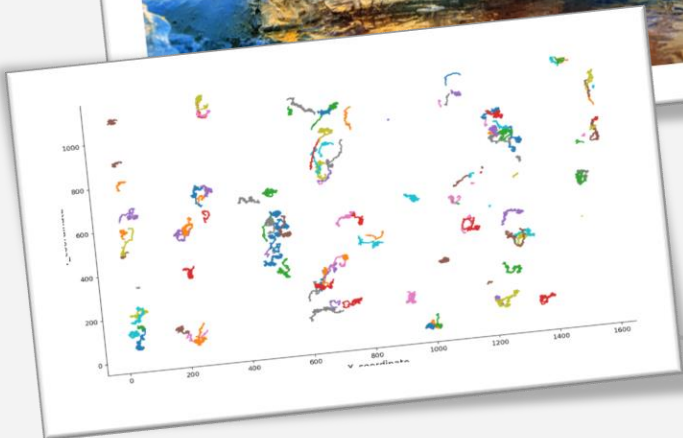
Возможности

- Снижение общей ошибки классификатора
- Выделение классов
- Подготовка мультиклассовой классификации
- Сохранение модели
- Подбор архитектуры сверточной нейронной сети (U-net)
- Подбор слоев нейронной сети, ядра свертки (Conv2D – 14x14)
- Работа над переобучением



Кейс 5

Изображения



Фотоника – трекинг частиц в суспензии

Задача

Создание универсальной платформы для прикладных специалистов при решении задач трекинга частиц в различных состояниях

Методы решения

Алгоритмы компьютерного зрения - сегментации и object detection.

Возможности

Алгоритмы компьютерного зрения - сегментации и object detection позволяют определить:

- наличие частиц суспензии
- их четкие границы
- координаты центра
- смещение относительно глобальных координат.

Определение объекта по звуку

Задача

- Распознавание объектов и мест по звуку.
- Определение объектов на зашумленных изображениях с помощью звуков.

Методы решения

- Объединение двух датасетов (звук и видео)
- Классификация (50 классов, 2000 образцов)
- Метод естественной синхронизации аудио и видео потоков (20 Тб)
- Подбор архитектуры нейронной сети SoundNet
- Обучение нейронной сети и тестирование

Перспективы применения

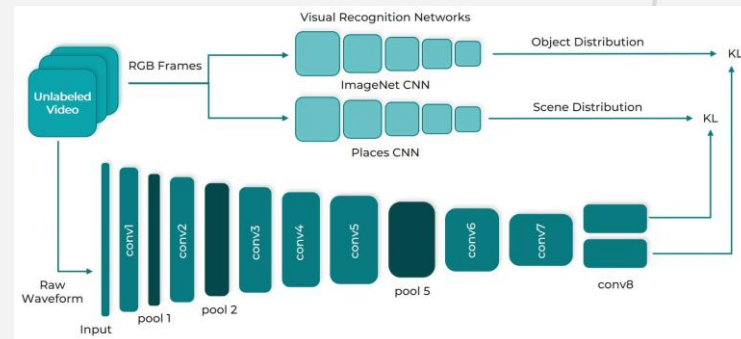
Возможность определить место действия по звукам — такая же важная задача, как и определение места по видеоматериалу.

Картинка с камеры зачастую может быть смазанной или не давать достаточно информации. Но если микрофон работает — робот уже сможет сориентироваться, где он находится.

Кейс 1

Звук

Архитектура SoundNet



	Accuracy on	
Method	ESC-50	ESC-10
SVM-MFCC	39,6%	67,5%
Convolutional Autoencoder	39,9%	74,3%
Random Forest	44,3%	72,7%
Piczak ConvNet	64,5%	81,0%
SoundNet	74,2%	92,2%
Human Performance	81,3%	95,7%

Кейс 2

Звук

Классификация музыкальных аудиофайлов

Задача

- Обработка базы данных аудио файлов и их классификация.
- Создание библиотеки с аудиофайлами разных жанров.

Методы решения

- Нейронные сети: полносвязные слои, сверточные слои
- Параметризация аудио – спектральный центроид, мел-кепстральные коэффициенты, частота цветности

