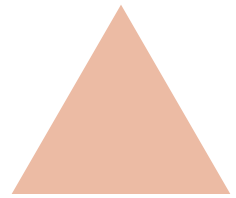




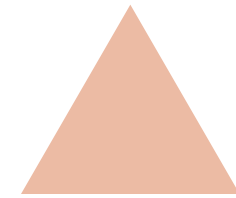
Artwell Analytics

Технология обработки «больших данных» (big data) из различных источников с использованием искусственного интеллекта

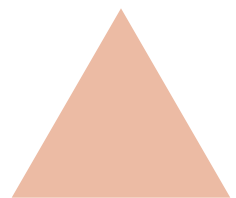
Проблематика рынка



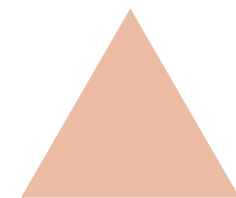
Сложность сбора и
качество обработки
больших данных



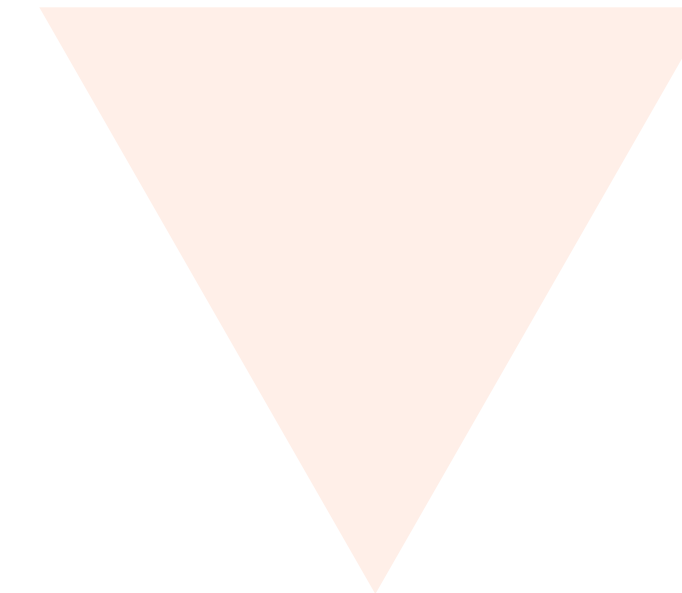
Отсутствие универсальных
инструментов по сбору и
обработке данных



Отсутствие коробочных
решений с ИИ, которые можно
быстро адаптировать под
конкретную задачу или бизнес

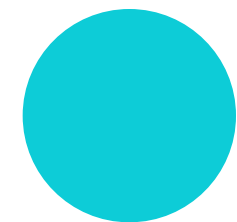


Высокая стоимость разработки
систем по работе с большими
данными

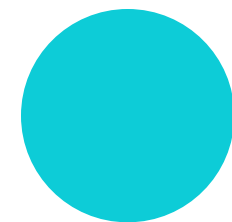


Решение

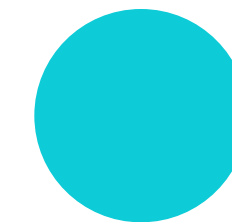
Разработан универсальный алгоритм, который позволяет:



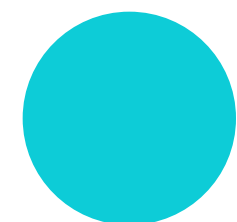
Обрабатывать большие данные и находить паттерны в информации



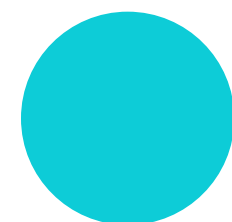
Собирать данные из различных источников: новостные сайты, чеки, блоги и т. д.



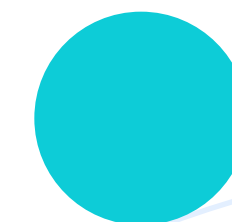
Автоматизировать процесс сбора при сокращении затрат на разработку не менее 40%.



Подключаться к различным системам заказчика быстро с помощью готовых скриптов и шаблонов



Обрабатывать данные с использованием ИИ с качественным показателем от 90%*



Универсальность технологии. Легкая адаптируемость под различные тематические области (Например, медицинская тематика, трейдинг, продуктовая и др.)

*Подтверждено классификацией чеков Росстата

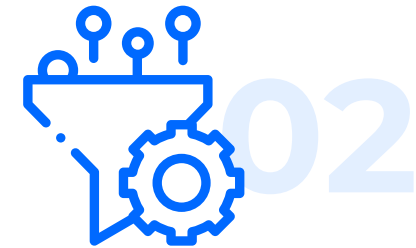
Чем мы отличаемся от остальных?

У нас есть **коробочное решение, которое экономит время и деньги** на разработку софта для работы с большими данными.

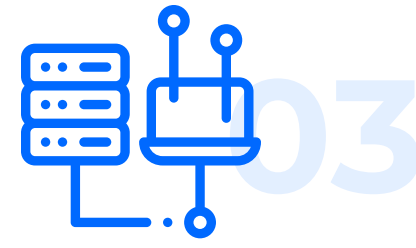
Возможности продукта



Система собирает данные из различных источников – базы данных, сайты, блоги, новостные ресурсы, интеграции (легко настраивается в административной панели)



Система классифицирует и распределяет данные по категориям/параметрам/бизнес-процессам (визуальная настройка)



Анализирует и визуализирует полученные данные (конструктор отчетов более 150 шаблонов).

Например, прогноз закупок для крупных ритейлеров или для успешной торговли на бирже.



Наш продукт – это приложение с возможностью настройки под конкретные задачи/процессы/функции



Точность распределения, обработки, классификации не менее 90%.

Подтверждено Росстатом.

Компонентная схема

Бизнес приложения (front-end): Vue.js

Сервисы (back-end): PyTorch, NumPy, Spring, Elastic Enterprise Search, Pandas

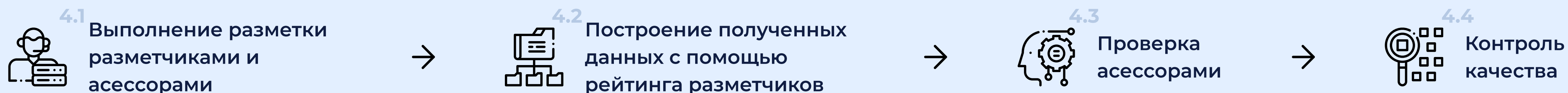
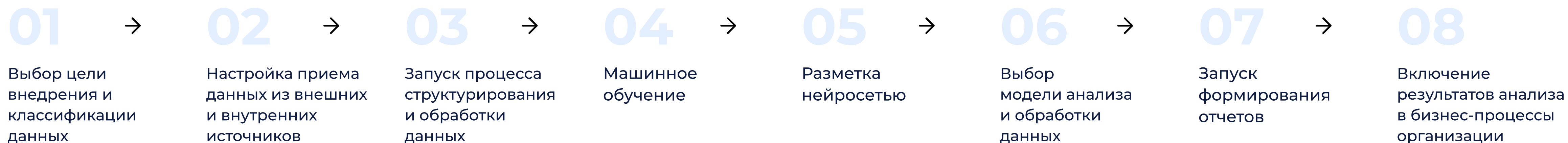
Хранение данных: Ceph, Spark SQL, Elastic Enterprise Search, Redis, Prometheus

Загрузка данных: Spark, Python

Существующий корпоративных ландшафт: Zabbix, Grafana

Установка Artwell Analytics

Сохраняем конфиденциальность данных



4.1.2

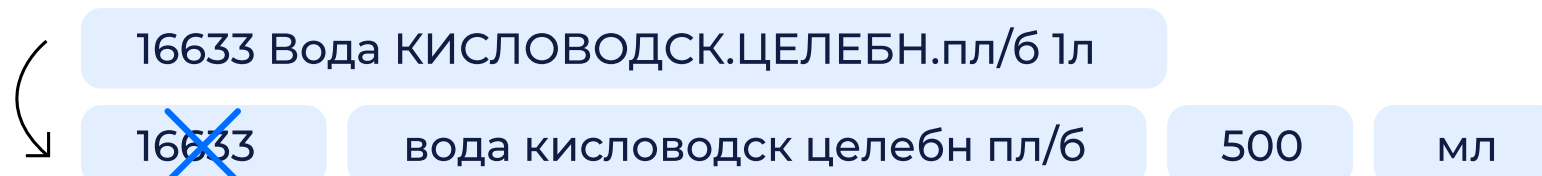


4.1.3



Обработка данных делится на 2 этапа

Токенизация – это разделение наименований чека на отдельные позиции, выделение единиц измерения и объема, приведение к нижнему регистру.



Первичная классификация – это унификация единиц измерения и разметка наименований из чеков по категориям.



Как работает нейросеть?

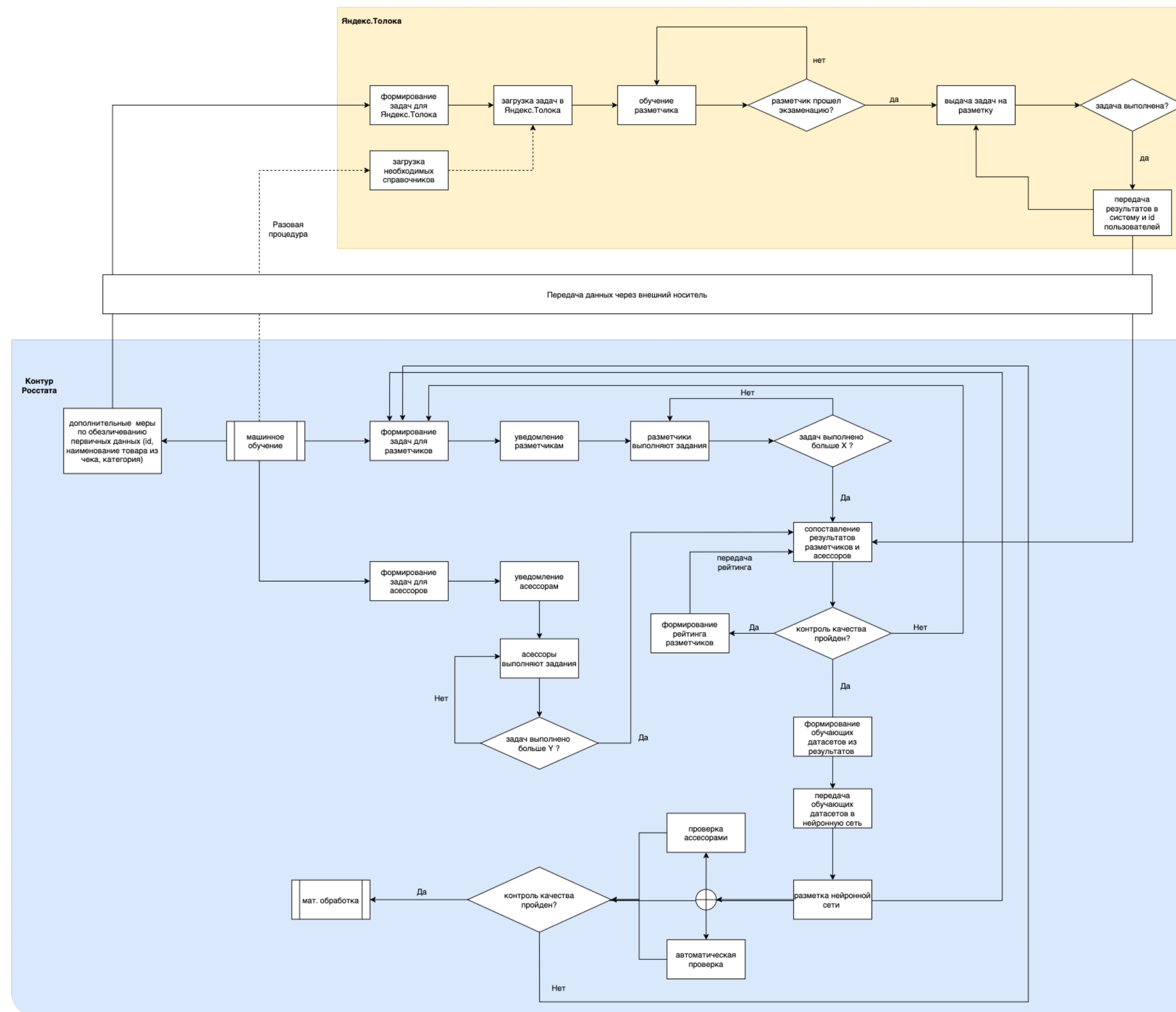
Обучение нейросети производится с помощью датасетов, подготовленных разметчиками. Пример разметки категории «Бараночные изделия»:

Наименование товара: Сушка мини от Каролины с маком 1 кг / Цена: 206

Верно Не верно Затрудняюсь ответить

Алгоритм обработки данных

Посмотреть полную схему



ИС БД Росстат



Описание задачи

Создать систему обработки и классификации больших данных для анализа статистики потребительских цен и статистики торговли.

Создать удобное web-приложение, для управления функциями математической обработки и контроля качества полученных данных.



Что было сделано

Внедрена система автоматизированного мониторинга данных фискальных чеков с применением алгоритмов big data.

Снижена нагрузка на отчитывающихся респондентов и сокращен ручной труд по сбору информации о ценовых котировках.

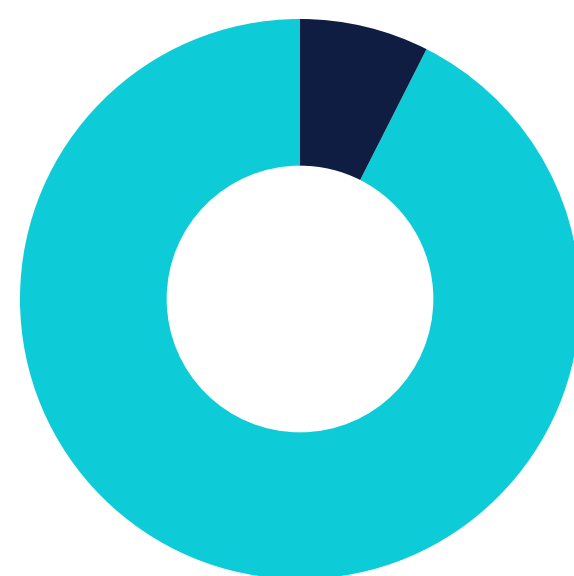
Применена технология больших данных с перспективой их использования в других направлениях статистики.

Разработан web-интерфейс для работы с ИС БД.

ИС БД Росстат

Результат

Внедрена в промышленную эксплуатацию система, позволяющая **проводить статистические исследования на основе данных ККТ**, получаемых на регулярной основе от ФНС России.



90%

ТОЧНОСТЬ
КЛАССИФИКАЦИИ
ДАННЫХ ИЗ ЧЕКОВ
НЕЙРОСТЕТЬЮ

ОБРАБОТКА ДАННЫХ | ИДЕНТИФИКАЦИЯ ДАННЫХ | КОНТРОЛЬ КАЧЕСТВА | РАЗМЕТКА ДАННЫХ | АДМИНИСТРИРОВАНИЕ

Главная / Идентификация данных

Категории

Поиск по наименованию и

Наименование
Колбаса сырокопченая
Бромгексин
Консервы мясные для де
Говядина
Таурин
Ксилометазолин (Галазо
Соки фруктовые
Колбаса вареная
Консервы мясные
Пельмени, манты, равио

К < 1 2

Фразы ИПЦ | Фразы ОКПД2 | Инструкция | Машинное обучение

Категория: Говядина | Код ИПЦ: 1504

Экспорт всех фраз | Импорт всех фраз

Поиск единиц измерения: Поиск | Присвоенные единицы измерения: Грамм x, Тонна x

Основная фраза: ГОВЯДИНА | Искключения из фразы: нарезка x, felix x, колбаски x, тушёная x, балык x, perfectfit x, perfect x, gourmet x, доширак x, д/собак x, соб x, собак x, тефтели x, шницель x, холодец x, барбекю x, по-восточному x, по-строгановски x, по-восточному x, по-еврейски x, по-мусульмански x, по-тайски x, картофелем x, wok-гуляш x, пастроми x, лангет x, кубики x, тушен x, +

Основная фраза: ГОВЯД | Искключения из фразы: нарезка x, felix x, колбаски x, тушёная x, балык x, perfectfit x, perfect x, по-еврейски x, по-мусульмански x, по-тайски x, картофелем x, wok-гуляш x, пастроми x, лангет x, кубики x, тушен x, +

Основная фраза: ГОВ | Искключения из фразы: чаппи x, корм x, шаурма x, вискас x, косточки x, бек x, гуляш x

Сохранить | Добавить фразу | Удалить | Отмена

© 2021 Информационная система по сбору больших данных. версия: 2.1 | Система | Документы | Форум | Те

Отличия от конкурентов

Наименование	Artwell ИИ	IBM SPSS Statistics	Stata США	IQPLATFORM РФ	M-Brain Intelligence Plaza Финляндия
Наличие модуля предсказательной аналитики	✓	✓	✓	—	—
Визуализация данных	✓	✓	✓	—	✓
Анализ больших данных	✓	✓	—	✓	—
Машинное обучение	✓	—	—	—	—
Потоковая аналитика	✓	—	—	—	✓
Коннекторы для источников данных	✓	✓	—	✓	✓
Интерактивная аналитическая обработка (OLAP)	✓	✓	—	—	✓
Индикация трендов и проблем	✓	✓	✓	—	✓
Панель администрирования	✓	—	—	✓	✓
Многопользовательский доступ (низкая плата на лицензию)	✓	—	✓	✓	✓
Наличие API	✓	—	—	✓	✓
Коробочное решение	✓	—	—	—	—
Возможность подключения необходимых модулей	✓	—	✓	—	✓

Бизнес-модель

Мы предлагаем полноценное коробочное решение, который позволяет подключаться к системе заказчика собирать и обрабатывать данные.

Алгоритм гибко настраивается нашей компанией под задачи клиента.

Модель монетизации



Автоматизация задач клиента по работе с “большими данными”
(оплата по контракту)



Поддержка и доработка внедренных систем
(ежемесячная оплата)



Покупка коробочного продукта без преднастройки с инструкциями

Модель коммерциализации

Мы предлагаем полноценную информационную систему с многочисленными настройками, которая решает задачи по сбору и обработке “больших данных” по различным бизнес-процессам.

Каналы продаж

- 01** Реклама продукта (GA/Я.Д/SMM/SEO)
- 02** Продажа постоянным клиентам: Mercedes-benz, Caterpillar, Royal Canin, LG и другие
- 03** Зарубежные партнеры-интеграторы
- 04** Международные выставки/конференции

Кому это подойдет?

- 01** Подойдет крупным ритейлерам для формирования закупок и анализа цен конкурентов.
- 02** В сфере недвижимости для анализа динамики цен и прогнозирования даты продажи.
- 03** Для торговли на бирже путем анализа новостей и рынков.

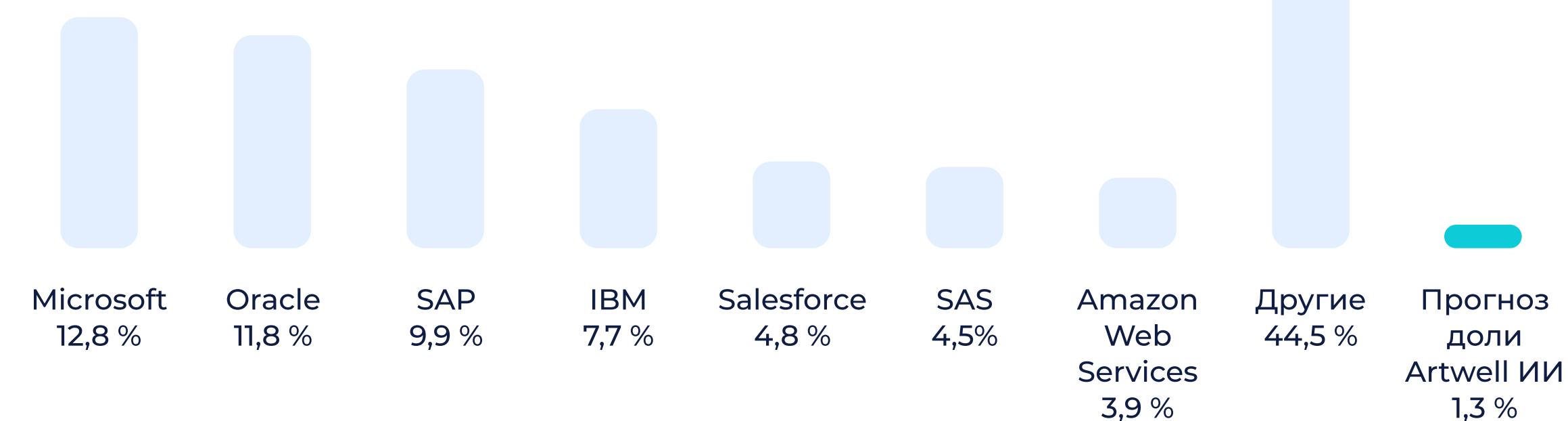
Коммерческий потенциал продукта

Наш продукт не уступает основным поставщикам ПО Big Data, что было подтверждено при внедрении продукта более чем десяти крупным компаниям. **Основные отличия нашей продукции: стоимость, простота внедрения, качественный легкоизменяемый код.** За счет более привлекательной стоимости и нашей репутации на международном рынке, **мы можем рассчитывать, как минимум, на 1% от общего объема рынка, что может составить 1,7 млрд.\$**

Доля рынка крупных поставщиков ПО больших данных в 2019

169 млрд \$

Объем рынка



10%

Ожидается, что глобальный рынок больших данных и аналитики будет расти со среднегодовым темпом равным 10%.

330 млрд \$

В прогнозируемом периоде к 2026 году он составит 330 миллиардов долларов США.

55,5%

Мировая доля лидеров рынка ПО Big Data составляет 55,5%, остальная доля рынка принадлежит менее известным компаниям.

Спасибо за внимание!

Теперь вопросы

