

ЦИТ

ЦИФРОВЫЕ
ИНДУСТРИАЛЬНЫЕ
ТЕХНОЛОГИИ

Система аналитики данных



Система аналитики данных

ЦИФРОВЫЕ
ИНДУСТРИАЛЬНЫЕ
ТЕХНОЛОГИИ

ЦИТ

Low-code система разработки решений для анализа данных и текстов, а также оптимизации и автоматизации бизнес-процессов. Аналитические сценарии строятся из последовательности функциональных узлов, являющихся операциями над данными.

ПЕРЕДОВАЯ АНАЛИТИКА «ИЗ КОРОБКИ»

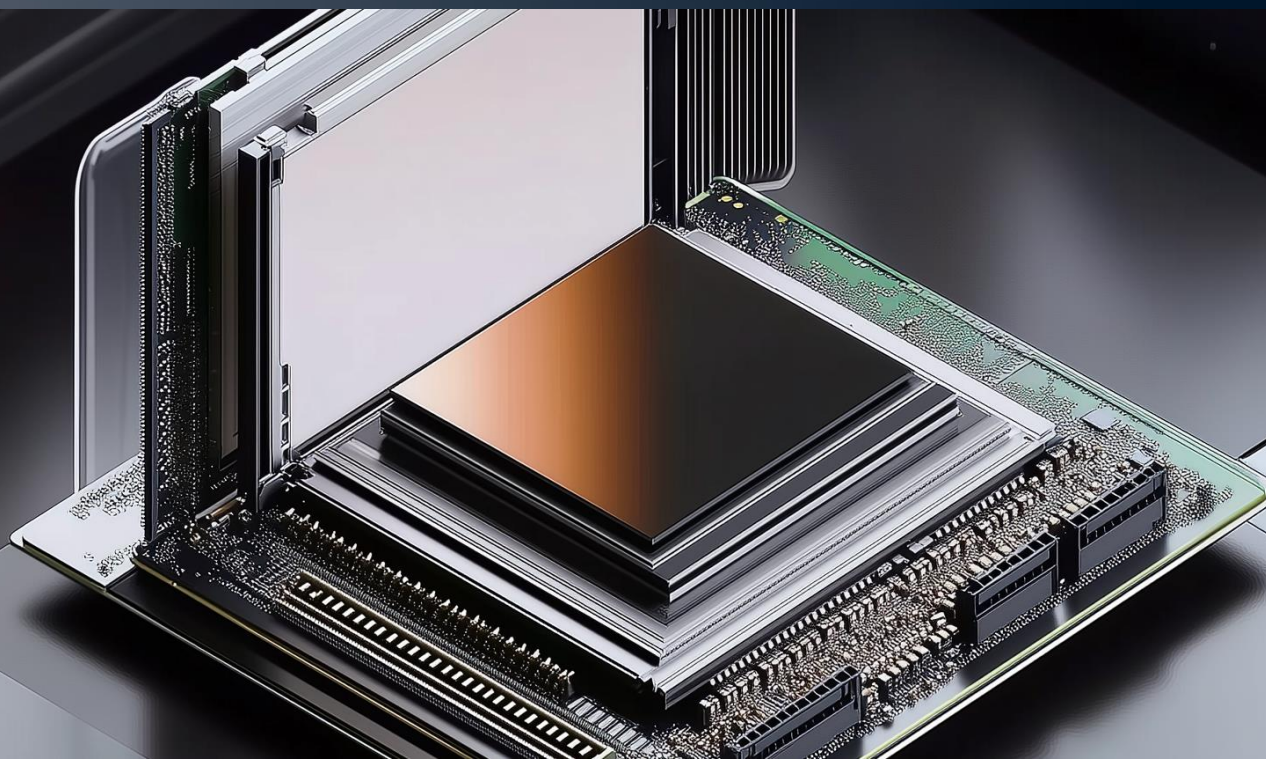
Встроены инструменты обработки данных, более 100 готовых алгоритмов машинного обучения и не имеющие конкурентных аналогов инструменты текстового анализа.

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ АНАЛИЗА

Платформа встраивается в информационную систему, может самостоятельно обрабатывать данные и выдавать результат без контроля со стороны пользователя.

ПОСТРОЕНИЕ ДАШБОРДОВ

Система включает собственные средства для визуализации результатов и построения интерактивных графических веб-отчётов.



Все этапы работы с данными

Это сквозная система, в которой имеются все инструменты для полного цикла работы с данными, что позволяет избежать необходимость эксплуатации и поддержки множества различных и слабо-совместимых программных решений от разных производителей.



Сбор данных из разных источников

- Загрузка данных из файлов, сканов документов, баз данных, документов, электронной почты, хранилищ данных



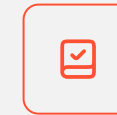
Подготовка данных

- Объединение данных, очистка, преобразование, агрегация, обогащение, поиск и удаление дубликатов, нормализация



Машинное обучение и текстовый анализ

- Извлечение из документов атрибутов и фактов, классификация документов, построение предиктивных моделей, анализ временных рядов



Визуализация и экспорт

- Построение интерактивных дашбордов, выгрузка результатов анализа в виде файлов, формирование регламентной отчетности по шаблонам MS Excel и MS Word

ИИ в Системе анализа данных

Система анализа данных обладает ИИ «под капотом», который способствует аналитике и разработке посредством системы, а также позволяет обучать собственные модели:

OCR

Собственный модуль распознавания текста и таблиц OCR на основе сверточных нейронных сетей

Speech to Text

Модуль транскрипции голосовых аудиозаписей в текст с собственным алгоритмом распознавания голосов (диаризация)

NLP

«Под капотом» у всего текстового анализа процедуры предварительной обработки текстов на основе собственных моделей:

- Определение языка текста
- Морфологический анализ
- Синтаксический анализ
- Семантический анализ
- Постаггинг (определение частей речи)

ML

Встроенные no-code алгоритмы для обучения моделей:

- Классический ML
- Кластеризация
- Ансамбли моделей
- Анализ временных рядов
- Анализ связей

GPT

Обучена собственная большая языковая модель, которая может быть развернута в контуре заказчика.

Через Систему обеспечивается доступ к модели GPT, а также другим языковым моделям, в том числе, GigaChat и ChatGPT

RAG

Retrieval Augmented Generation или Генерация с дополненной выборкой – метод работы с текстом, который основан на схеме вопросов и ответов. Вопрос от пользователя обрабатывается ИИ на основе загруженных данных. Далее языковая модель формулирует ответ в соответствии с полученными вводными данными

Варианты импортозамещения

Система аналитики данных внесена в Единый реестр российских программ для электронных вычислительных машин и баз данных.
Офисы разработки находятся в Москве и Чебоксарах.
Системные компоненты и функциональные модули собственной разработки.

1. Extract/Transform/Load

- SAP BW ETL, IBM DataStage, Alteryx, Tableau Prep, Pentaho Kettle
-

2. Машинное обучение

- SAS Enterprise miner, IBM Watson (SPSS Modeler), RapidMiner
-

3. Текстовый анализ

- SAS Text Analytics, KNIME, IBM Watson
-

4. Бизнес-аналитика

- Qlik, Power BI, Tableau, SAS Visual Analytics
-

Возможности Системы аналитики данных

Система позволяет решать широкий набор задач и автоматизировать бизнес-процессы в компаниях различных отраслей посредством применения алгоритмов машинного обучения (ML) и работы с естественным языком (NLP - Natural Language Processing).

Анализ документов

Сбор, сведение и анализ разноформатных отчетных данных (электронные таблицы, текстовые документы, сканы)

Выявление рисков

Построение предиктивных моделей, выявление аномалий и закономерностей

Кадровый HR анализ

Анализ должностных инструкций и положений, извлечение функций/задач/обязанностей, создание конструктора должностных инструкций

Анализ данных из Интернет

Контроль состояния контрагентов через анализ новостных сводок, изучение мнения потребителей, конкурентная разведка

Авто-классификация

Автоматическая обработка поступающей корреспонденции, определение типов документов и установление их взаимосвязей

Аналитика call-центра

Систематизация записей из контактных центров и анализ полученных данных

Кейс: повышение эффективности аудита и экспертно-аналитических мероприятий в Счетной палате РФ

ЦИФРОВЫЕ
ИНДУСТРИАЛЬНЫЕ
ТЕХНОЛОГИИ

ЦИТ

ЗАКАЗЧИК: Счетная палата РФ

Проблема: Инспекторы Счетной палаты в своей деятельности сталкивались с проблемой крайне высоких сроков сбора и объединения данных из множества разнообразных источников. Кроме того отчетные данные из регионов поступают в ведомство в различных форматах, как в виде электронных таблиц, так и в виде текстовых документов, а зачастую в неанализируемом формате растровых изображений документов (сканов). Рядовой инспектор не обладает навыками программирования на Python или VBA и написания SQL-запросов, что приводит к необходимости привлечения дополнительных специалистов для предварительной подготовки и очистки больших объемов разноформатных данных. Также в силу применения исключительно классических инструментов работы с данными, таких как MS Excel, достигается невысокая точность результатов анализа.

Задача: Повысить точность и глубину анализа данных без программирования с помощью low-code инструмента.

Решение

- Благодаря применению Системы аналитики данных сотрудникам Счетной палаты РФ удалось значительно увеличить объем и глубину обработки данных, которые используются при экспертно-аналитических мероприятиях. Также появилась возможность автоматически обрабатывать новые виды данных - неструктурированные текстовые документы, которые составляют существенную долю информации, поступающей в Счетную палату. Кроме того, применение методов Data Science, входящих в функциональность Системы аналитики данных, повысило точность анализа. Одновременно с ростом качества результатов существенно сократились затраты времени и труда на проведение аналитических процедур.

Результаты и эффект

- Наличие широкого набора коннекторов в Системе позволило с высокой скоростью осуществлять сбор данных для анализа из большинства имеющихся в организации информационных систем;
- Возможность одновременной обработки, очистки и структурирования десятков тысяч документов самим владельцем данных исключила необходимость привлечения профессиональных программистов, что сократило затраты, ускорило процесс и снизило количество ошибок;
- Сокращение ручного труда при анализе и значительный рост объема выборки привели к повышению точности и репрезентативности результатов;
- Сокращение издержек стало возможно благодаря тому, что загрузку, подготовку, анализ данных и разработку итоговых интерактивных отчетов и другие работы теперь можно проводить в одной системе.

20-70%

Разгрузка профессиональных разработчиков, системных и бизнес-аналитиков после внедрения

x2

Увеличение вовлеченности бизнес-пользователей в аналитические проекты и процедуры разработки решений

☰ **Экономия времени разработки:** ~8 часов на проект вместо 20 человеко-дней

🔍 **Экономия человеческих ресурсов:** 1 разработчик вместо 4. Среднестатистический аналитический проект разрабатывается без привлечения специалистов в программировании или аналитике самим «владельцем данных»

📁 **Повторяемость анализа:** ∞ кол-во использований. Разработанный один раз аналитический проект может использоваться регулярно

ЦИТ

ЦИФРОВЫЕ
ИНДУСТРИАЛЬНЫЕ
ТЕХНОЛОГИИ

Благодарим за внимание!



t.me/cit_gov



cit.gov.ru

