

KNIME Analytics Platform

Платформа для анализа данных и машинного обучения

Команда:

Илья Морозов

Владислав Куликов

Александр Грибков



Что такое KNIME?

KNIME — Konstanz Information Miner

- Open-source платформа для анализа данных
- Визуальное программирование без кода
- Более 300 коннекторов к источникам данных
- 6 лет подряд — лидер Gartner Magic Quadrant

История:

- **2004** — разработка в Университете Констанца (Германия)
- **2006** — первое коммерческое применение
- **Сегодня** — миллионы пользователей по всему миру



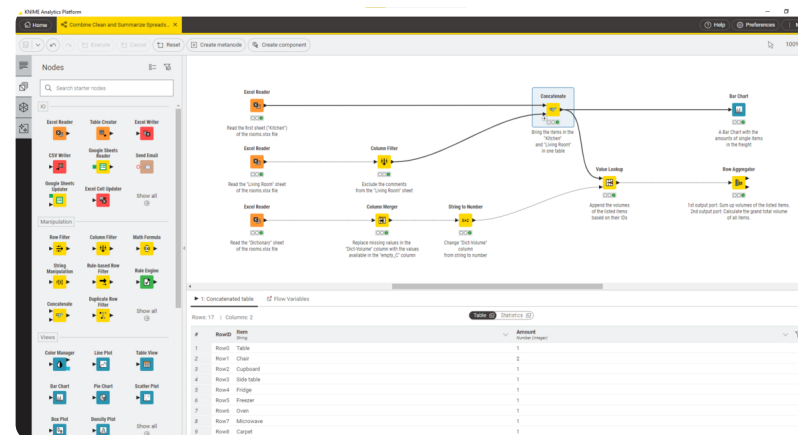
Основные задачи KNIME

- **ETL** — извлечение, трансформация, загрузка данных
- **Data Mining** — поиск паттернов в данных
- **Machine Learning** — обучение предиктивных моделей
- **Визуализация** — интерактивные графики и дашборды
- **Автоматизация** — повторяемые аналитические процессы
- **Отчётность** — генерация отчётов



Интерфейс KNIME Analytics Platform

| Область | Назначение |
|----------------------|------------------------------------|
| Репозиторий узлов | Библиотека всех доступных операций |
| Рабочее пространство | Область построения workflow |
| Настройки узла | Конфигурация выбранного узла |
| Консоль | Логи выполнения и результаты |



Принцип: Drag & Drop

Принцип работы

Визуальное программирование в KNIME

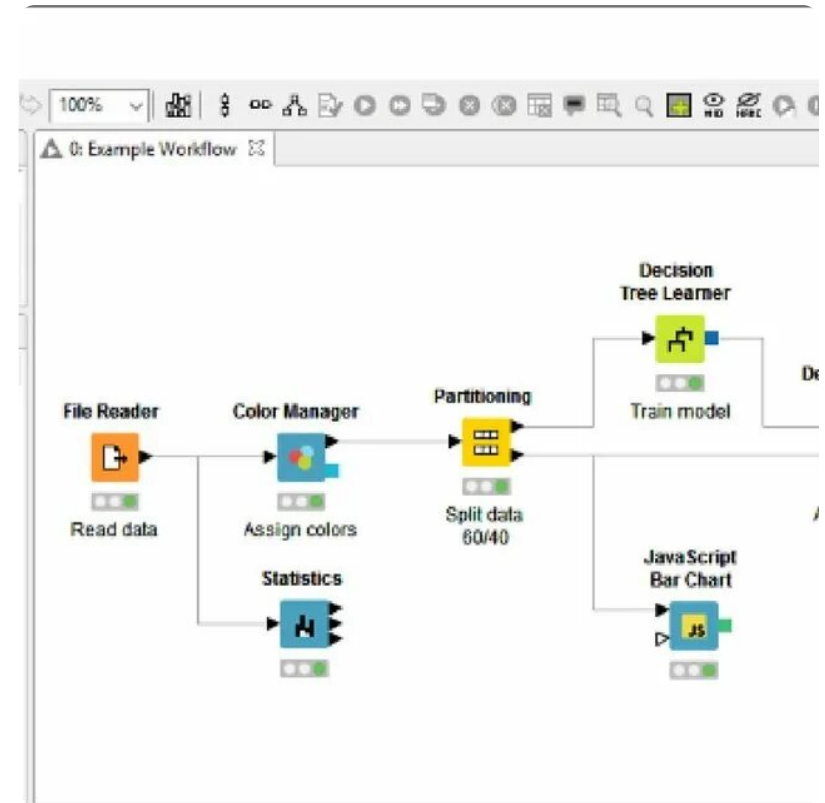
Источник данных → Обработка → Анализ →
Результат

Workflow (рабочий процесс):

- **Узлы (Nodes)** — отдельные операции
- **Соединения** — поток данных между узлами
- **Выполнение** — последовательное или параллельное

Статусы узлов:

- **Красный** — не настроен / ошибка
- **Желтый** — выполняется
- **Зеленый** — успешно завершён



Узлы — строительные блоки KNIME

Структура узла:

- Входные порты — получение данных
- Выходные порты — передача результатов
- Настройки — конфигурация операции

Категории узлов:

| Категория | Примеры |
|--------------|-------------------------------------|
| IO | File Reader, Database Connector |
| Manipulation | Filter, Joiner, GroupBy |
| Analytics | Decision Tree, K-Means, Regression |
| Views | Scatter Plot, Histogram, Table View |

Источники данных в KNIME

Файлы:

- CSV, Excel, JSON, XML, Parquet

Базы данных:

- MySQL, PostgreSQL, Oracle, SQL Server, MongoDB

Облачные сервисы:

- AWS S3, Azure Blob, Google Cloud Storage

Big Data:

- Hadoop, Apache Spark, Hive

API:

- REST API, Web Services, SOAP

Более 300 коннекторов из коробки



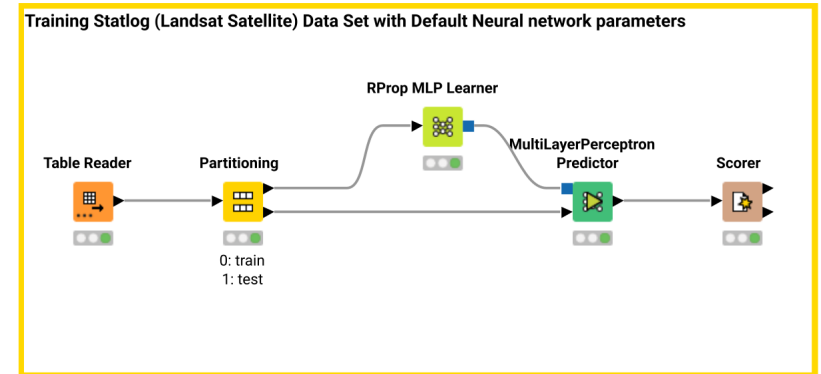
Machine Learning в KNIME

Поддерживаемые задачи:

- Классификация (Decision Tree, Random Forest, SVM)
- Регрессия (Linear, Polynomial, Gradient Boosting)
- Кластеризация (K-Means, DBSCAN, Hierarchical)
- Deep Learning (нейронные сети, CNN, RNN)

Полный цикл ML:

1. Загрузка и подготовка данных
2. Разделение на train/test
3. Обучение модели
4. Оценка качества
5. Применение к новым данным



Интеграция с языками программирования

| Язык | Возможности |
|------------|---|
| Python | pandas, scikit-learn, TensorFlow, PyTorch |
| R | ggplot2, caret, dplyr, tidyverse |
| Java | Собственные узлы, расширения |
| JavaScript | Интерактивные визуализации |

Как это работает:

- Python Script — выполнение Python-кода внутри workflow
- R Snippet — встраивание R-скриптов
- Двусторонний обмен данными между KNIME и скриптами

Почему выбирают KNIME

Open Source

- Бесплатная базовая версия
- Открытый исходный код
- Нет vendor lock-in

No-Code подход

- Визуальное программирование
- Доступно для не-программистов
- Быстрое прототипирование

Сообщество

- KNIME Hub — тысячи готовых workflow
- Активный форум
- Регулярные обновления

Признание индустрии

- 6 лет — лидер Gartner Magic Quadrant
- Используется в Fortune 500

KNIME в фармацевтике

Задачи

- Анализ клинических испытаний
- Поиск побочных эффектов
- Предсказание эффективности

Workflow

Данные → Анализ → ML-модель →
Результаты

Результат

- 3-5x ускорение исследований
- Снижение затрат на анализ
- Воспроизводимость экспериментов

KNIME в финансовом секторе

Задачи:

- Кредитный скоринг
- Выявление мошенничества
- Оценка рисков

Пример: Кредитный скоринг

Заявка → Обогащение данных → Скоринг → Решение

Преимущества:

- Автоматизация решений по кредитам
- Обнаружение подозрительных транзакций в реальном времени
- Соответствие регуляторным требованиям (аудит workflow)

KNIME в маркетинге и CRM

Ключевые задачи:

- Сегментация клиентов
- Прогнозирование оттока
- Персонализированные предложения

Пример:

RFM-анализ → K-Means → Сегменты

Результаты:

- Конверсия +20-30%
- Снижение оттока
- Оптимизация бюджета

Где изучать KNIME

KNIME Hub

hub.knime.com

- Тысячи готовых workflow
- Компоненты и расширения

Официальная документация

docs.knime.com

- Руководство пользователя
- Примеры и туториалы

Бесплатные курсы

knime.com/learning

- Self-paced курсы
- Сертификация

Сообщество

forum.knime.com

- Активная поддержка
- Ответы на вопросы

Спасибо за внимание!

