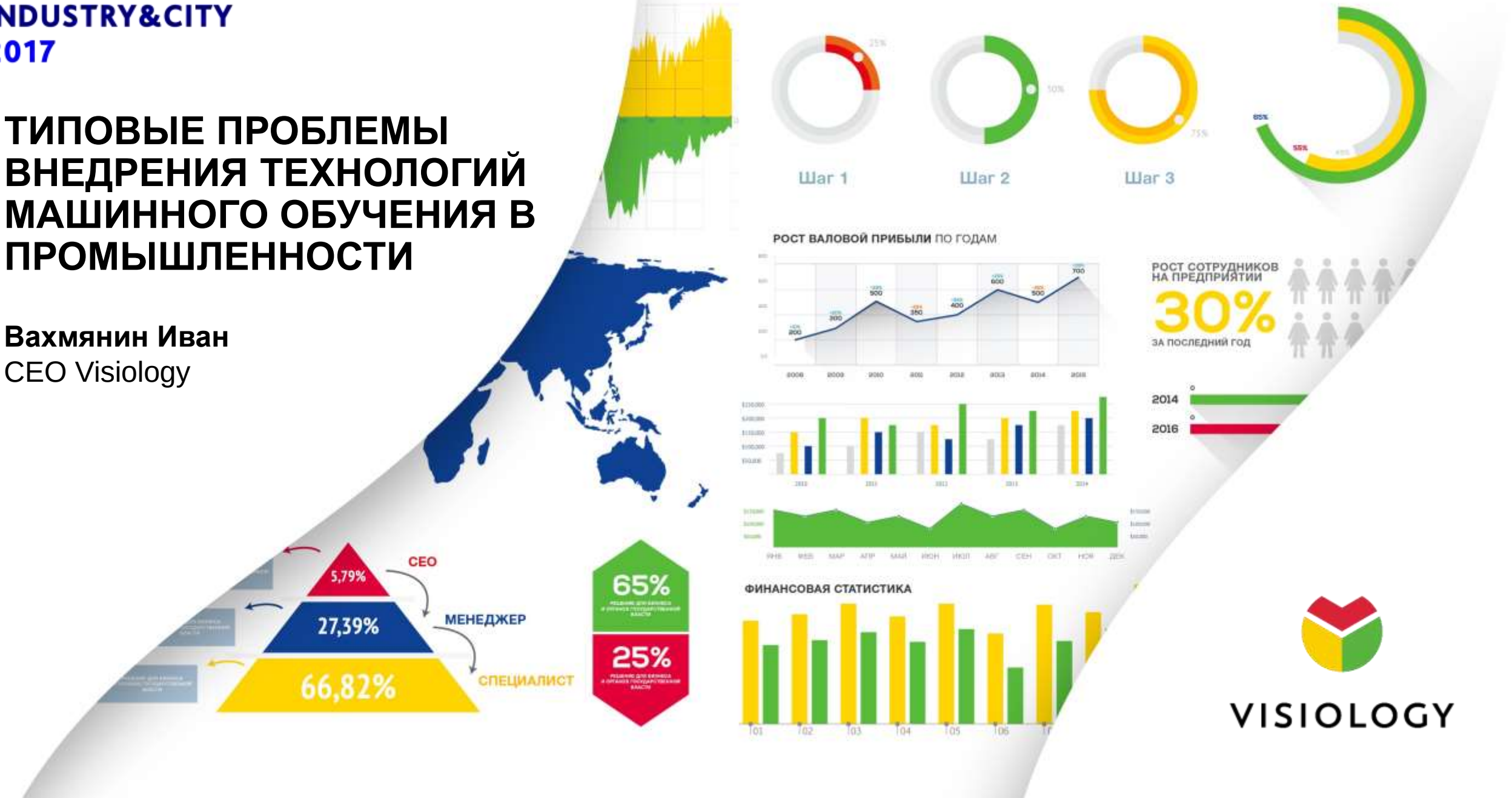
SMART INDUSTRY&CITY 2017

ТИПОВЫЕ ПРОБЛЕМЫ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ В ПРОМЫШЛЕННОСТИ

Вахмянин Иван
CEO Visiology



VISIOLOGY

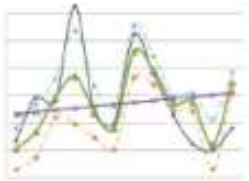
Аналитические решения Visiology

Аналитическая платформа Visiology

Современная платформа класса Business Intelligence для быстрого построения производительных аналитических систем для коммерческих и государственных организаций.

1. In-Memory база данных ViQube для быстрого выполнения аналитических запросов
2. Встроенная система сбора данных через настраиваемые веб-формы
3. Мощная и расширяемая система визуализации и поддержка видеостен
4. Поддержка интеграции с технологиями Big Data и Data Science
5. Полностью российская разработка

Собственный штат специалистов по Data Science

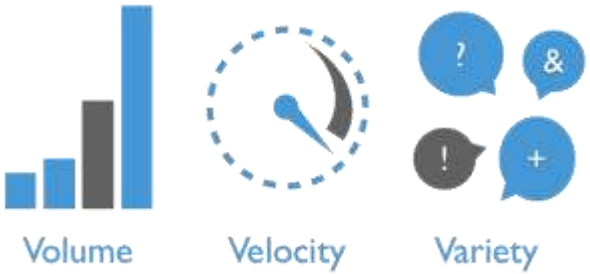


$$r_{pbi} = \frac{\frac{1}{n}\sum^{n} x^{*} - p_i \bar{x}}{S_x \sqrt{p_i q_i}}$$

Специалисты Visiology имеют опыт работы с большими данными и имеют необходимую математическую и техническую подготовку (R, Python, Spark и т.д.), чтобы разрабатывать и внедрять решения продвинутой аналитики на основе технологий машинного обучения (Machine Learning)

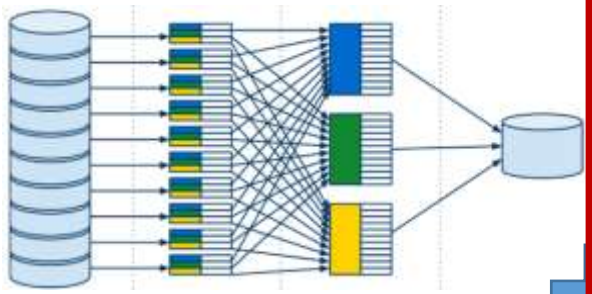
VISIOLOGY

Компоненты решения Big Data



Данные, которые слишком велики для обычных средств хранения и слишком сложны для традиционных средств аналитики

Big Data Tools

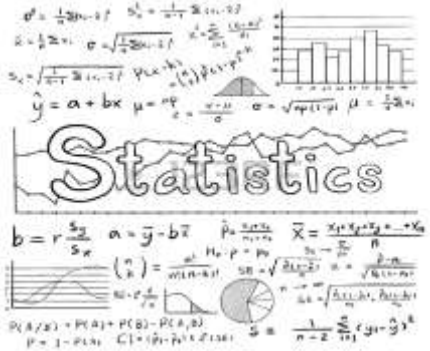


Инструменты для сбора, хранения и анализа данных, распределенных по кластерам серверов

Hadoop Apache Spark

Инженеры Big Data VISIOLOGY

Data Science



Математические и алгоритмические методы, оптимизированные для эффективного выявления сложных закономерностей

R Spark MLlib

Специалисты Data Science VISIOLOGY

Business Intelligence

Визуализация и публикация данных для простого использования конечными бизнес-пользователями

Аналитическая платформа VISIOLOGY

Data Science - как это работает?

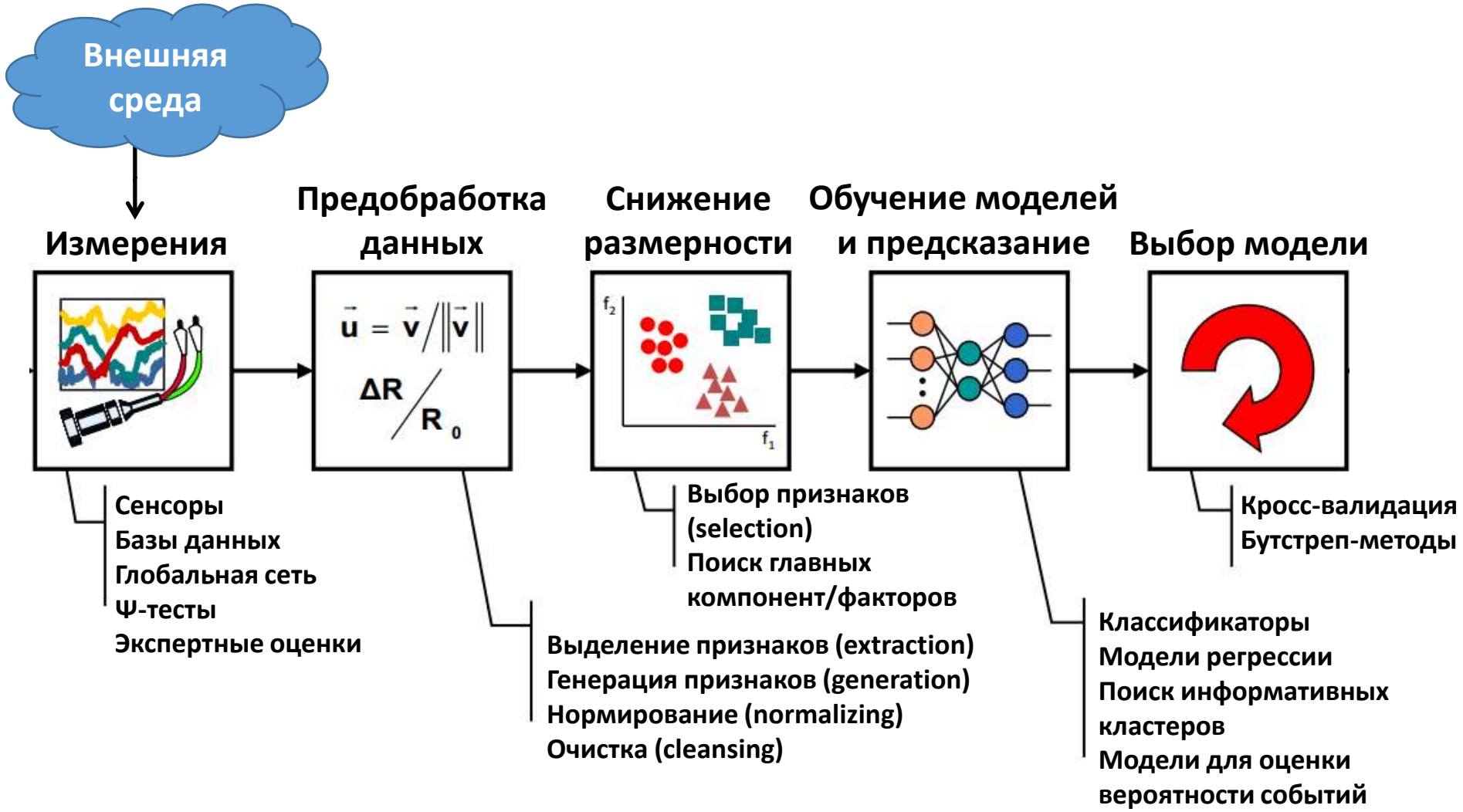


Иллюстрация взята с http://research.cs.tamu.edu/prism/lectures/pr/ © Ricardo Gutierrez-Osuna | CSE@TAMU

VISIOLOGY

ПРОБЛЕМА №1

Плохая формулировка бизнес-цели

НЕ документированная
НЕ измеримая
НЕ пересчитанная в экономический эффект
НЕ согласованная с топ-менеджментом

VISIOLOGY

Понимание бизнеса

Необходимо:

1. Измеримые критерии успешности
2. Оценка экономического эффекта

Пример формулировки бизнес-цели

Бизнес-целью проекта является снижение отходов производства цеха металлообработки П1.2 на 1,5% в течение тестового периода (1 месяц) по отношению к среднему значению за 2016 год.

Показатель отходов производства рассчитывается методом материального баланса по методике РД-1234 с учетом исключения периодов пуска-остановки и ремонта.

Экономический эффект составит 30 млн. руб. в год на протяжении 5 лет за счет снижения затрат на закупку сырья и сокращения персонала, занимающегося уборкой отходов (2 единицы). Эффект утвержден экономическим отделом АО «Хххххх».

ПРОБЛЕМА №2

Недостаточное понимание данных

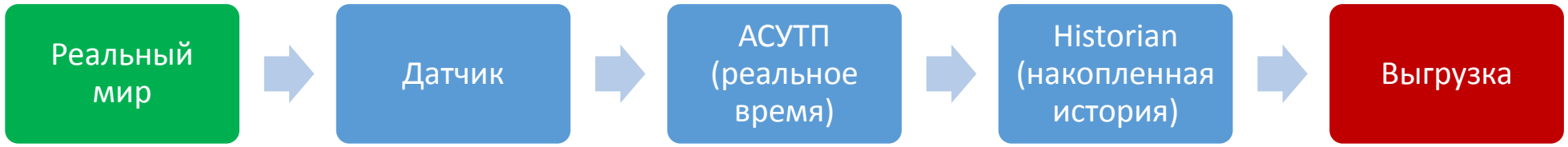
«Итак, датасет есть, приступим к моделированию!»

Плохой Data Scientist

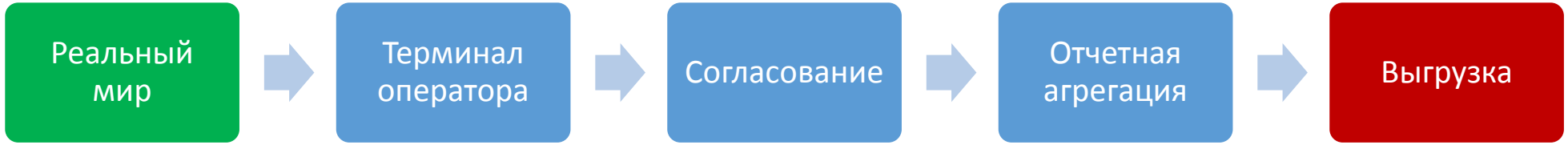
VISIOLOGY

Понимание данных
Цепочка преобразований

Машинный сбор данных (IoT, IIoT транзакционные системы)

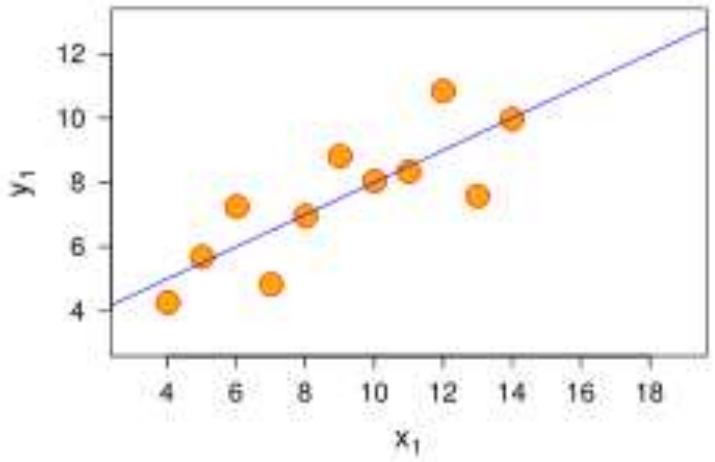


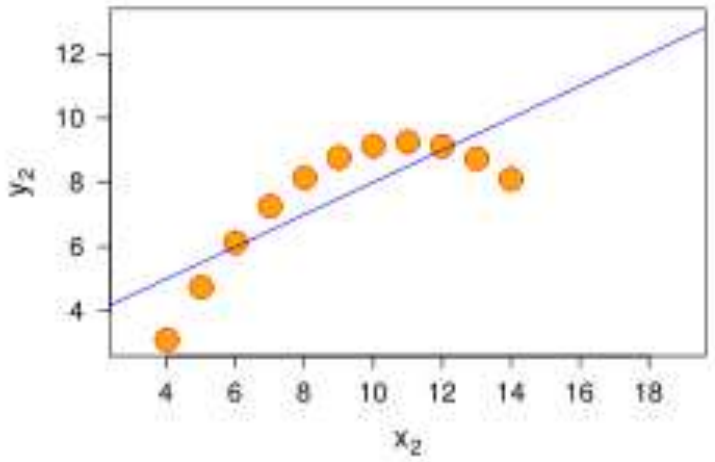
Ручной ввод данных (отчетные данных, системы с оператором)

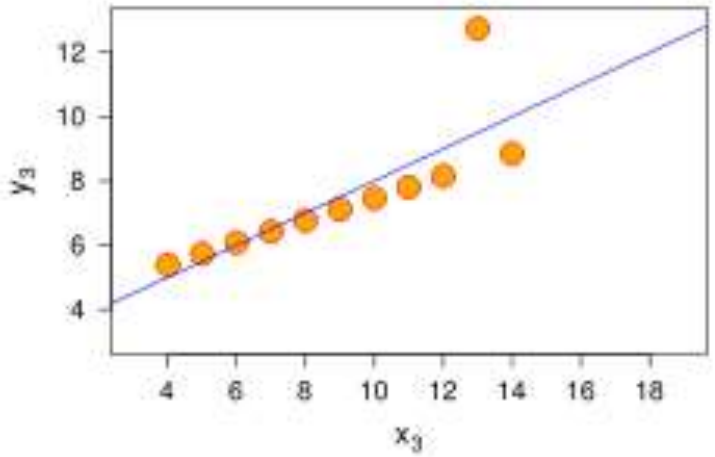


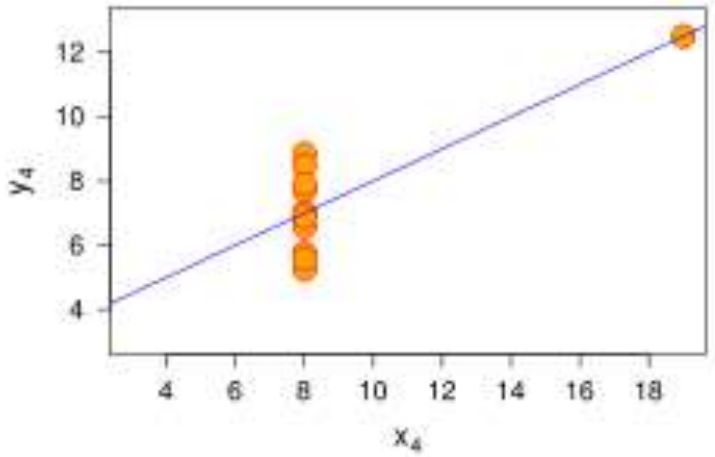
VISIOLOGY

Почему понимание данных критически важно?









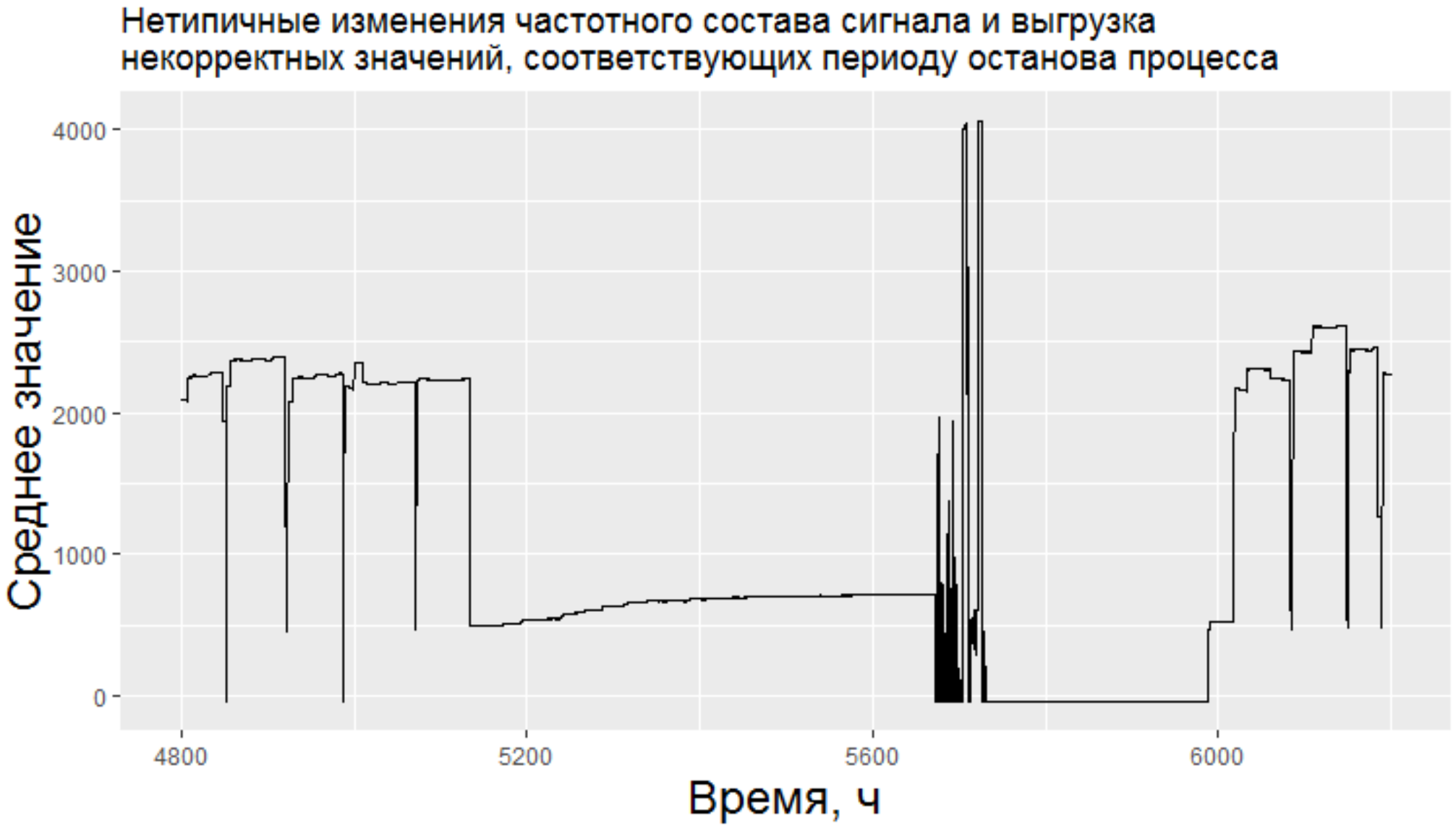
Характеристика	Значение
Среднее значение переменной x	9.0
Дисперсия переменной x	10,0
Среднее значение переменной y	7,5
Дисперсия переменной y	3,75
Корреляция между переменными x и y	0,816
Прямая линейной регрессии	$y = 3 + 0{,}5x$

Квартет Энскомба. https://ru.wikipedia.org/wiki/Квартет_Энскомба

VISIOLOGY

Машинный сбор данных

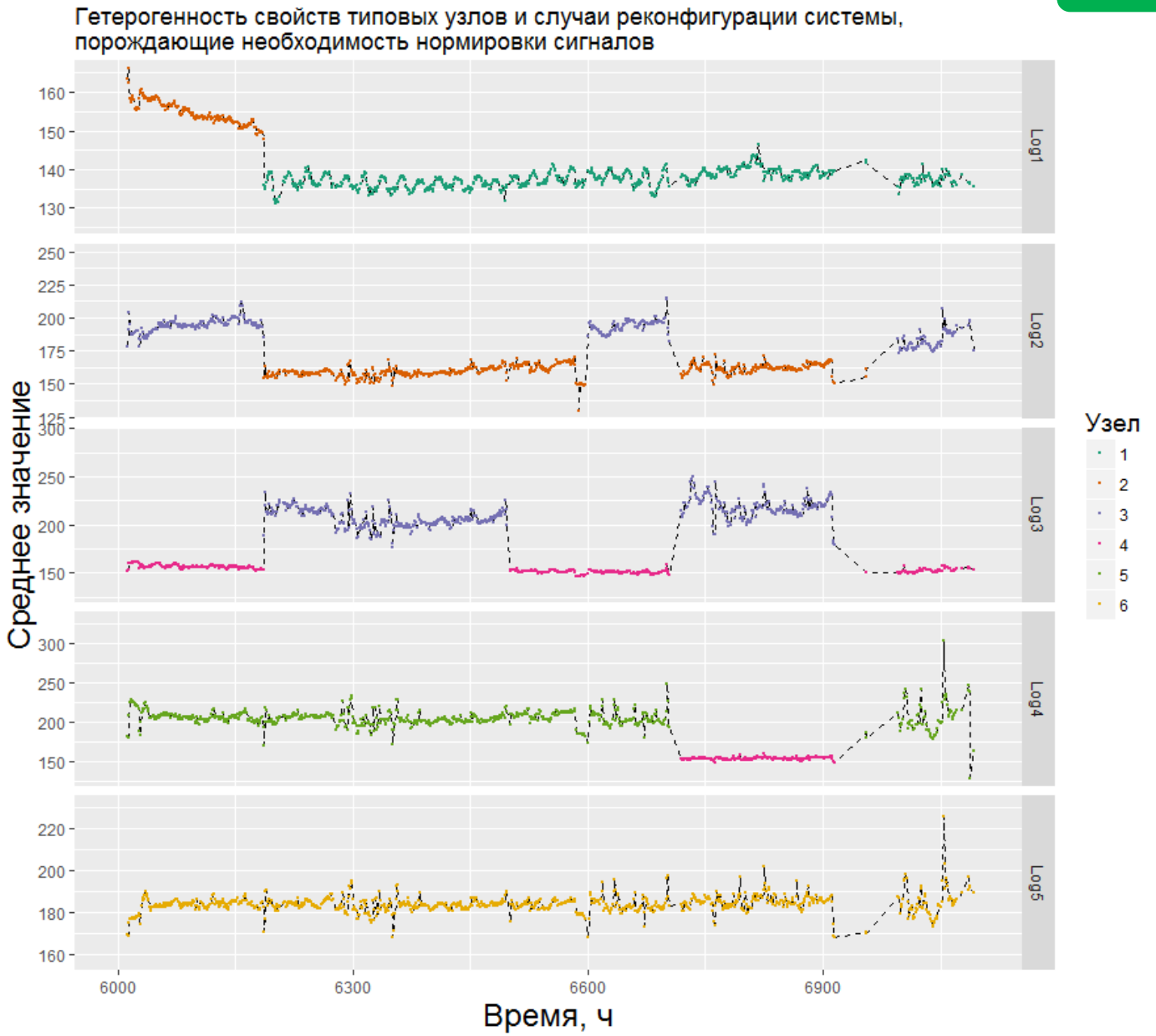
Реальный мир → Датчик



VISIOLOGY

Машинный сбор данных

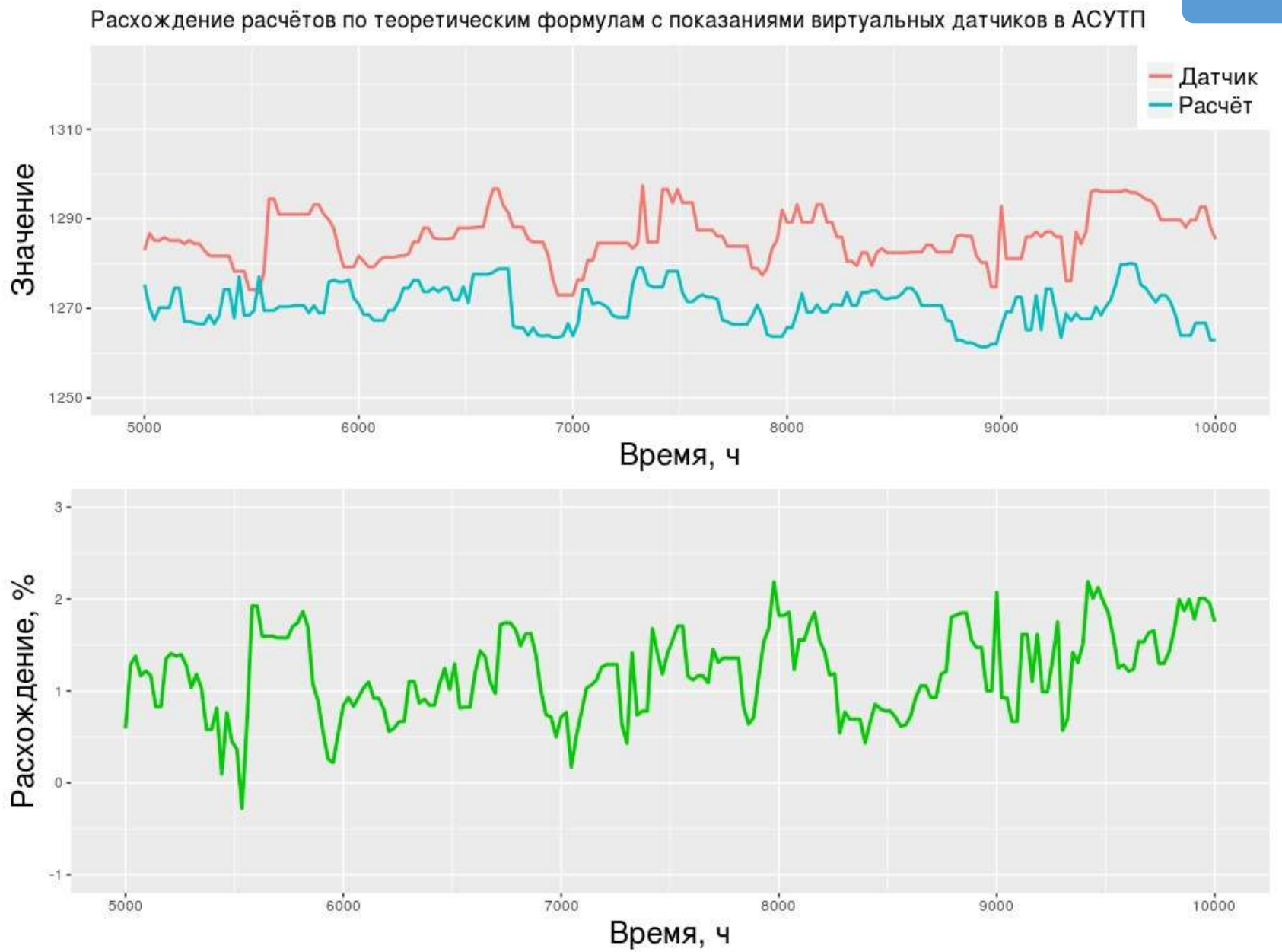
Реальный мир → Датчик



VISIOLOGY

Машинный сбор данных

Датчик → АСУТП



VISIOLOGY

ОШИБКА #3
НЕДОСТАТОЧНАЯ РАБОТА АНАЛИТИКОВ НА ПРОИЗВОДСТВЕ

Обследование за 3-4 консультации, пару поездок и в остальное время работа с датасетами и документацией

VISIOLOGY

РЕАЛЬНОСТЬ

Регулярные 1-2 недели на производстве в начале проекта и контакт с технологическими экспертами залог:

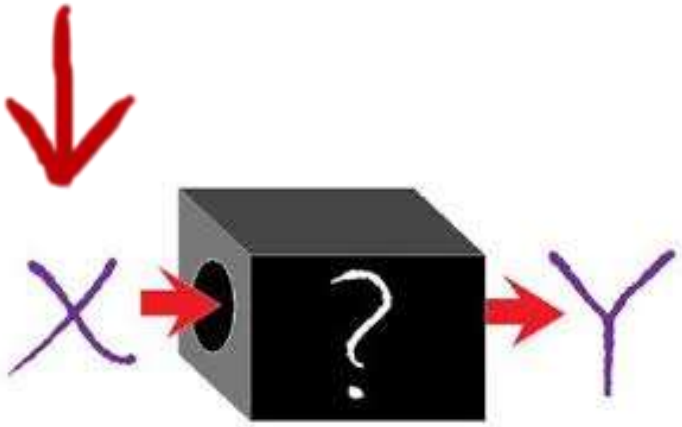
- ✓ качественного обследования
- ✓ Качественной разметки датасета и отбора обучающих выборок
- ✓ понимания процесса
- ✓ правильной постановки задачи и цели
- ✓ Корректировки и проверки модели в процессе R&D

VISIOLOGY

ОШИБКА #4
ПОСТРОЕНИЕ МОДЕЛЕЙ ТОЛЬКО МЕТОДОМ ЧЕРНОГО ЯЩИКА

«Большие данные»
на производстве

Очищенные данные



OGY

РЕАЛЬНОСТЬ

Очищенные данные

Пригодные для моделирования данные

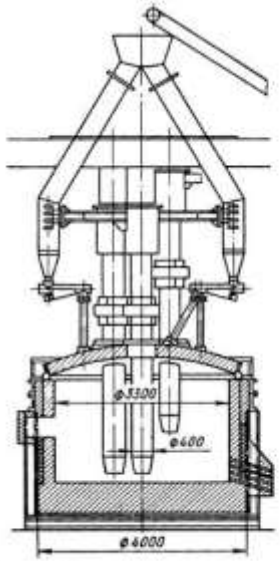


Рис. 106. Схема опытной рудотермической печи фирмы "Крупп"

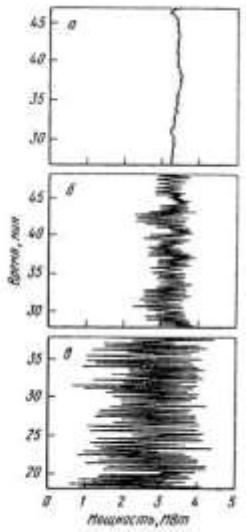
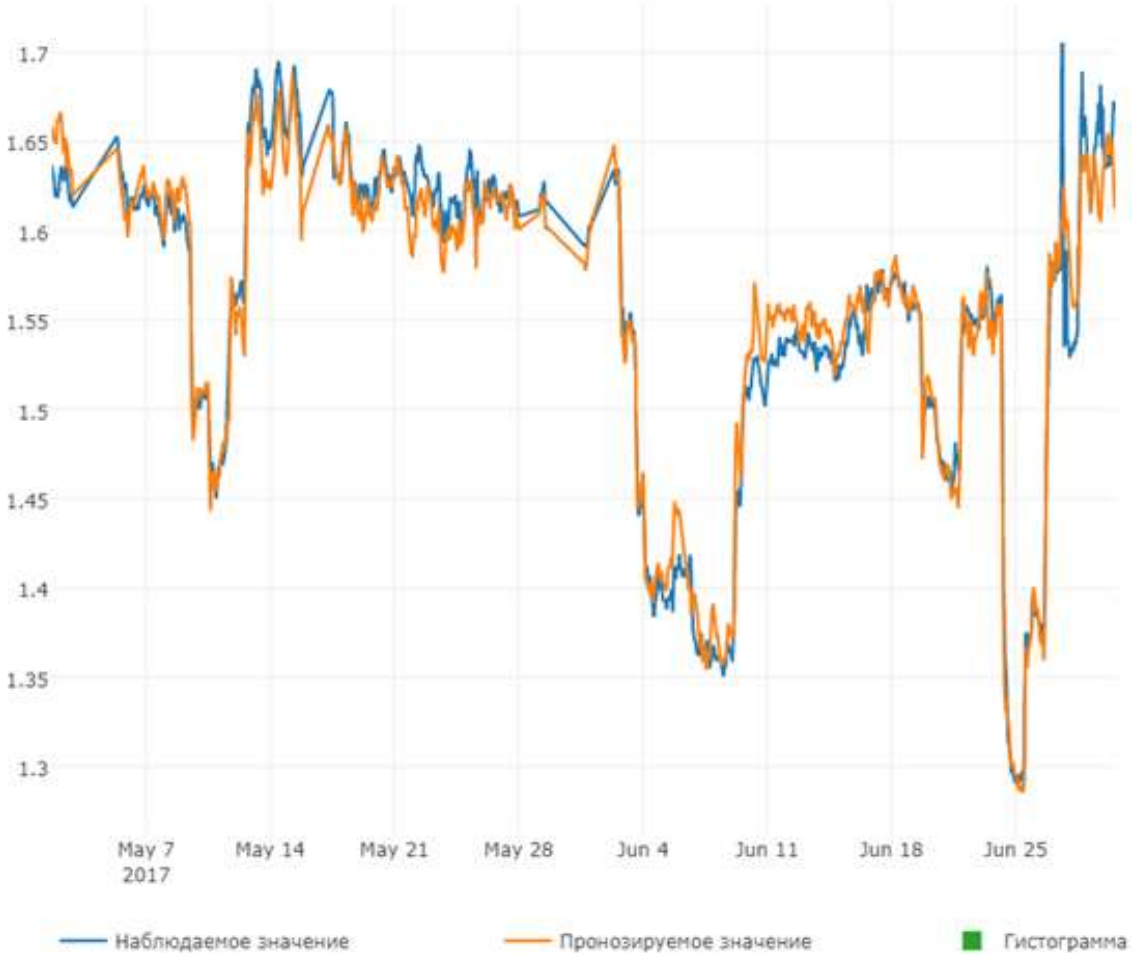


Рис. 107. Потребляемая мощность на опытных плавках при различном составе шлака

- ✓ После очистки и преобразования остается слишком мало данных для моделей «черного ящика»
- ✓ Сложные зависимости между параметрами, которые модели не улавливают – часть производства «черный ящик»
- ✓ Возможно какая-то часть важных данных и измерений отсутствует или не измеряется вовсе
- ✓ Особенности оборудования, оснастки и расходного материала, расчеты – секретное «ноу-хау» производителя, которые он тщательно охраняет даже под NDA
- ✓ Необходимо понимать производство и расчеты для сокращения размерности моделей

VISIOLOGY

ОШИБКА #5
ДОВЕРИЕ СЛЕПЫМ ТЕСТАМ ПРИ ОЦЕНКЕ КАЧЕСТВА МОДЕЛИ



Решая задачи с данными реального времени, нельзя полагаться на успешные результаты слепого теста

VISIOLOGY

РЕАЛЬНОСТЬ

Модель должна проходить несколько экспериментальных проверок в процессе разработки

- ✓ На эксперименте ошибки становятся в 2 раза заметнее и их величина и значимость растет
- ✓ Данные меняются в режиме реального времени, имеют внешние возмущения (воздействия) затрудняющие получение чистого результата
- ✓ В непрерывных процессах сложно заставить оператора ТП выйти из зоны комфорта привычных действий

VISIOLOGY

ОШИБКА #6
НЕДОСТАТОЧНОЕ ВОВЛЕЧЕНИЕ ИСПОЛНИТЕЛЕЙ И ПОЛЬЗОВАТЕЛЕЙ

4 причины почему важно плотно работать с будущими пользователями модели:

1. Получить нужную информацию
2. Получить нужную ценность и функциональность (см.ошибку #1)
3. Повысить доверие к результату
4. Вовлечь пользователей и обеспечить дальнейшее использования модели пользователями в рабочем процессе

VISIOLOGY

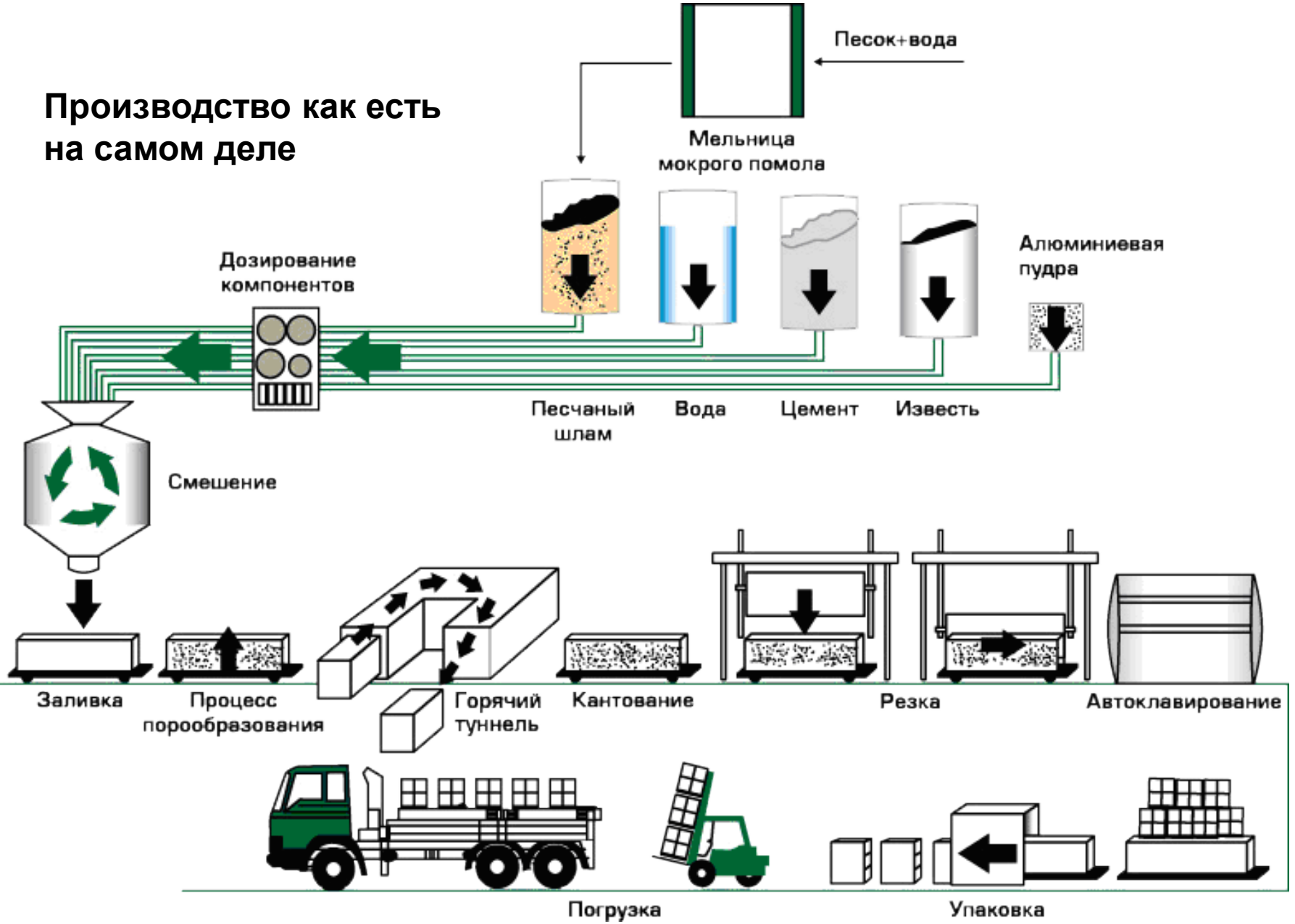
ОШИБКА #7
НЕДОСТАТОЧНАЯ ОЦЕНКА ИЗМЕНЧИВОСТИ ПРОЦЕССА

Производство как нам кажется в начале

VISIOLOGY

РЕАЛЬНОСТЬ

Производство как есть на самом деле



4 причины постоянной проверки модели на адекватность:

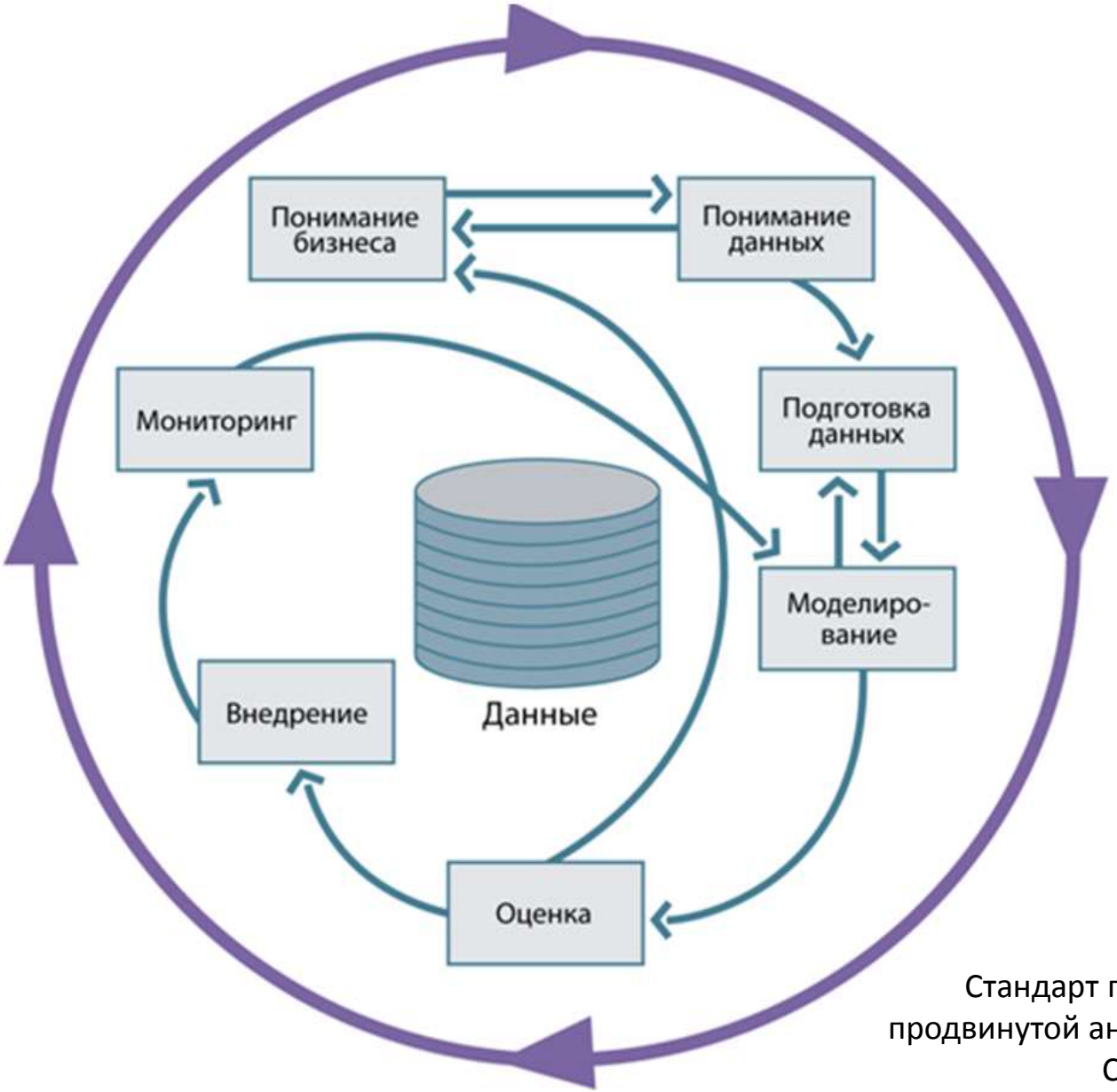
1. постоянные внутренние и внешние возмущения на процесс могут менять его параметры
2. необходимо четко понимать границы применимости модели (допущения, граничные условия)
3. сложная мат. модель может выдавать неправильные результаты
4. высокая стоимость ошибки

VISIOLOGY

Data Science - как это работает?

Data Science – это набор методик и практик для решения задач продвинутого анализа данных, в том числе:

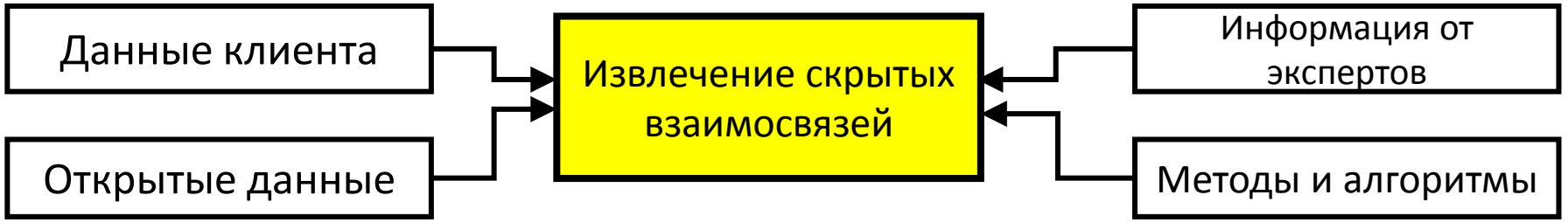
1. Подготовка (очистка)
2. Моделирование с применением методов **машинного обучения**
3. Оценка и верификация
4. Визуализация



Стандарт процесса продвинутой аналитики CRISP-DM

Типовой план проекта

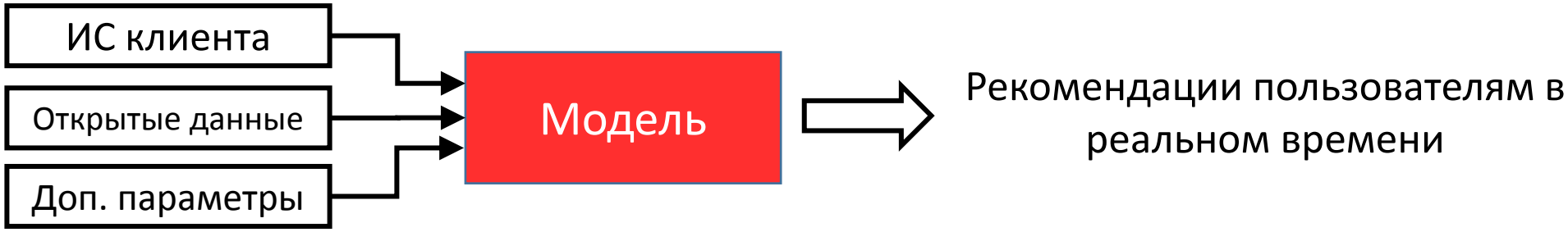
1. Разведочный анализ и предварительное моделирование **(~3 месяца)**



2. Разработка математических моделей и интеграция с источниками данных **(1-4 месяца)**

3. Опытная эксплуатация и корректировка модели **(2-4 месяца)**

4. Внедрение в промышленную эксплуатацию, поддержка.



VISIOLOGY

www.visiology.su

VISIOLOGY