Искусственный интеллект и данные

Мифы, рифы

Часть 2

Юрий Сирота

2022-02-10
CNews. Искусственный Интеллект 2022

Добро пожаловать, плохие новости!

Столкновение с неожиданностью – признак неадекватной картины мира

В.Тарасов

Introduction

В Machine Learning, Data Science, AI «ломанулись» практически все организации, скупая технологии и людей

Но успешных кейсов не так много, 5-10% успеха. Почему?

Причины неудач:

1. На волне хайпа не понимают целей
2. Организационные проблемы
3. Кадровые проблемы

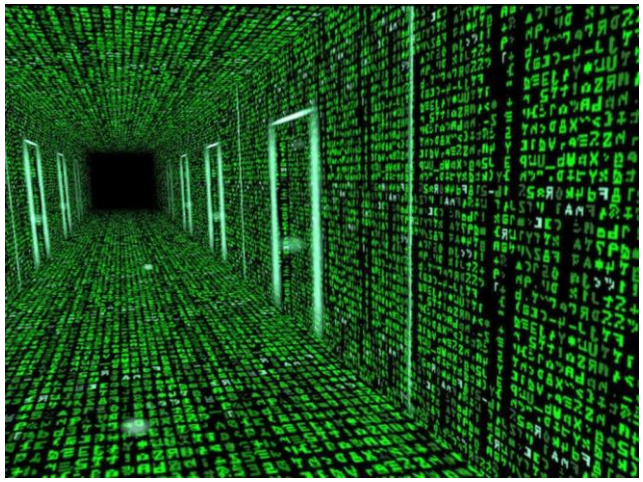
Хайп

Хайп (от англ. hype — «шумиха») — агрессивная и навязчивая реклама, целью которой является формирование предпочтений потребителя. Название её происходит от слова, означающего надувательство, обман или трюк для привлечения внимания. Wiki

Большие данные = Большая ценность?

Ценность данных не измеряется их масштабом

big data VALUE ≠ big data



Данные имеют ценность тогда, когда

- их проанализировали
- на основе анализа совершают управленческие действия

Слабый ИИ

- не имеет разума
- ориентирован на решение прикладных задач
- усиливает возможности человека в решении узких задач
- не функционирует без человеческого контроля

На сегодняшний день создан только слабый ИИ

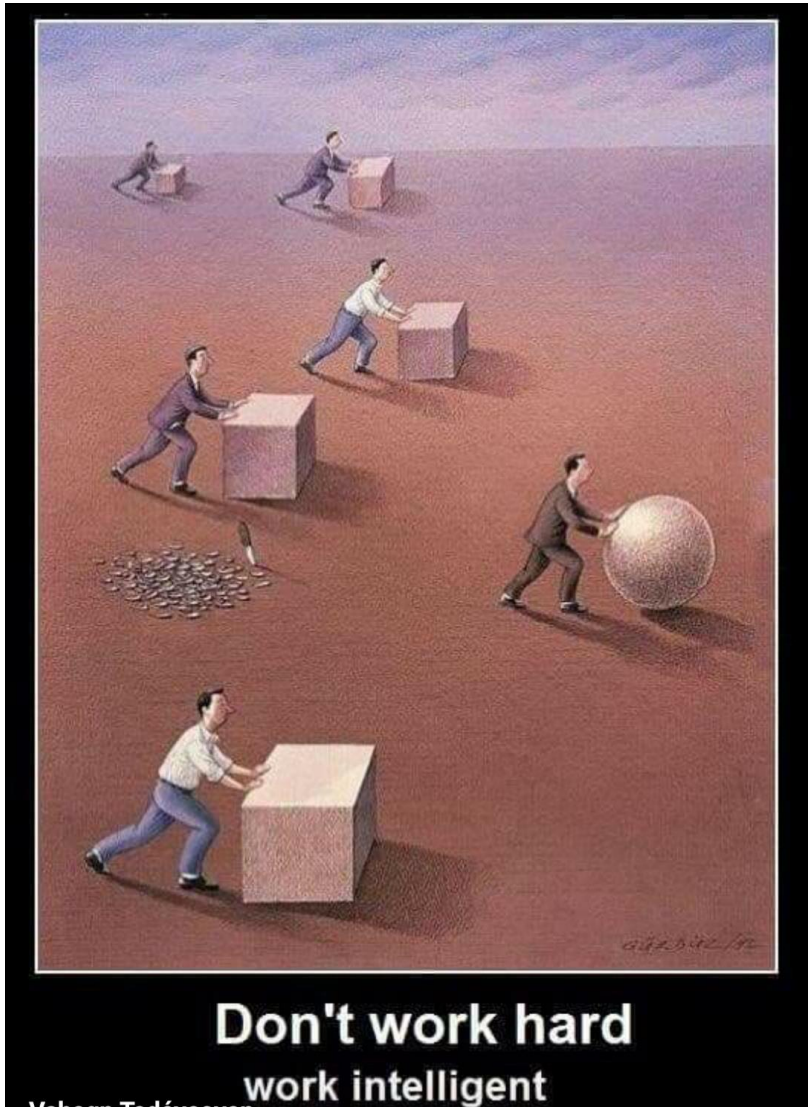
Decision Intelligence: работать много или работать умно?

Конкуренция возрастает, привычным масштабированием: открытие дополнительных офисов, найм клиентских менеджеров - не достичь значимого эффекта.

Работать умно помогает (не заменяет) направление **Decision Intelligence**:

- Математика, оптимизация, моделирование
- Data Science (AI)
- Process mining

Подходы «катать квадратное, носить круглое» уходят в прошлое



Система принятия решения. Пример фондового рынка

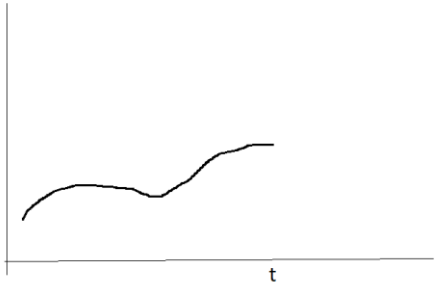
- Пусть актив Y наблюдается до момента времени t. Последнее известное значение цены Y(t)
- Дан прогноз Y(t+1) на период вперед. Для прогноза использовали кофейную гущу, астрологию, фундаментальный анализ или математику
- Если Y(t+1) > Y(t), покупаем актив, иначе продаем (пренебрегли стоимостью денег, комиссиями и т.п.).
- Наше действия сейчас (покупка/продажа) = функция от прогнозного значения

Action today = function(Forecast tomorrow)

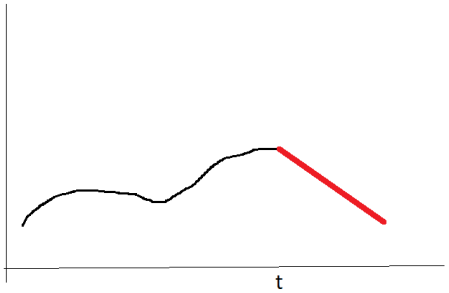
- Прогноз – математическое ожидание будущего значения. Чем меньше дисперсия прогноза, т.е. большая уверенность в точности прогноза, тем большее число акций покупается/продается.
- **Вывод**: Прогноз – основа принятия решения!!!

Система принятия решения. Общий случай

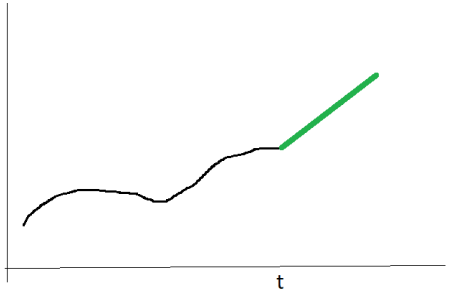
Пусть сущность Y наблюдается до момента времени t. Для нас чем больше Y, тем лучше. Что нам делать в момент t?



Допустим, удалось получить зависимость Y от X1, X2, X3: Y=f(X1,X2,X3). И получили прогноз на завтра (красная линия), который нас огорчил. Значит уже сейчас мы предпринимаем меры: организационные, кадровые, бизнесовые, чтобы улучшить тенденцию. Здравый смысл нам в помощь.



Однако помимо здравого смысла можно применить туже закономерность Y=f(X1,X2,X3), которую используем как рекомендательную систему (предписательная математика). Закономерность говорит, что нужно увеличить X1, уменьшить X2 чтобы прогнозный Y снова рос.



Цифровой двойник. Данные

Цифровой двойник – набор данных и связей между данными. Двойник позволяет осуществлять «виртуальный» эксперимент над объектом, а значит выбирать наилучшую траекторию развития и способ управления.

Данные: набор правильно упорядоченных по времени и характеристикам многомерные структуры хранения. К примеру, физический клиент характеризуется

- X1 – транзакция по дебету,
- X2 – транзакция по кредиту,
- X3 – наличие кредита…
- Xn и т.п. и обязательно ось времени.

Если структуры в DWH созданы неправильно, цифровой двойник не реализуем.

Цифровой двойник. Связи между данными

Связи между данными: трех типов для финорганизации

- **Финансовая модель** - самая общая и главная. Оперирует производными показателями от абсолютных данных и содержит множество ручных смысловых настроек. Некоторые элементы финансовой модели связаны математически, а встроенный математический оптимизатор (симулятор) позволяет найти наилучшую траекторию

- **Математические связи**. Применимы для моделирования абсолютных (количественных, объемных) показателей. Математические модели характеристик X1, X2, X3,…, Xn являются частью финансовой модели

- **Процессная связь**. Данные не связаны формулами, но описаны методы перетекания одних показателей в другие с помощью условных операторов и последовательностью преобразований. В процессной связи тоже применима оптимизационная математика, позволяющая найти наиболее эффективный процесс.

Все говорят: «Данные – новая нефть» ?!

Нефть	Данные
Извлечение нефти из недр. Создание нефтеналивного хранилища	Извлечение данных из источников. Создание DWH: очистить данные и скомпоновать витрины
Переработать нефть в бензин. Нефтеперерабатывающий завод	Переработать данные в информацию и знания. Математический отдел
Сформировать парк автомобилей и штат квалифицированных водителей	Сформировать новые бизнес-процессы в соответствии с новыми знаниями. Участники: владельцы процесса , математики, IT
Создание ценности: перевозка из пункта А в пункт В	Создание ценности: новый процесс эффективнее прежнего

Данные → Знания → Процесс

Создание ценности требует компетенции в
(1) Хранилищах
(2) Математике
(3) Бизнес-процессе

Многие организации стартуют с DWH и называют «управлением данными» или «data-driven»?
DWH – необходимое, но недостаточное условие, само по себе – расходная составляющая

Как оценивать Chief Data Officer?

Почему CDO чаще ограничиваются технологической составляющей?
Ответ: проще создать DWH , чем менять процесс. Математические рекомендации и изменение процессов накладывает ответственность!

Критерии оценки CDO на рынке?

- Количество людей в data-office
- Количество технологий

Оба критерия приводят к необоснованному росту затрат на раздутый штат,

затрат на зоопарк избыточных технологий и его поддержку

Как должно быть?

Чистый экономический эффект = приращения выручки (сокращение затрат) –

затраты на содержание дата офиса (включая железо, soft и штат).

PS

Почти всегда решение математическое лучше экспертного решения

О ловушке self-service. Замысел

Идея Self-Service очень заманчива – действительно, «запихиваешь» все в ПО и оно само все делает: self-service BI, self-service AI. Чудеса!!!

Идея Self-service нравится всем:

- Вендору ПО, который сможет продать лицензии большему числу пользователей
- Менеджеры децентрализуют задачи и не несут персональную ответственность за результат, в противоположность централизованной структуре
- Работникам, которым не self-service недоступен по причине нехватки квалификации и они ощущают свою причастность к великому, имея чудо-инструмент. Им теперь не нужно передавать свои задачи профессионалам и терять свою значимость
- Всей организации, т.к. «экспертиза развивается на местах»

О ловушке self-service. Наблюдаемое

Однако, что часто наблюдаем на практике?

- Сотрудники, пользующие self-service, задачи чуть более сложные продолжают делегировать профессионалам в центре компетенции, как то подготовить dash-board в BI.
- Интерпретировать результаты, выдаваемые волшебным ПО, не могут, т.к. не понимают алгоритмы, использованные под капотом. Словно функция ПРОГНОЗ в экселе: натравил формулу на выбранный диапазон клеток и получил значение прогноза…Чем не чудеса? Не понимаем как, но эксель же почитал! Но BI же посчитал!

Self-service – как очки у обезьяны в известной басне:

- профессионалы пользуются полноценным ПО
- любители свалят неудачи на чудо ПО.

Организация несет затраты на ПО

Инновации. Следствия

- нарушаются обычаи
- изменяются процессы
- права инноватора становятся больше
- чьи-то права ущемляются
- передел зон ответственности
- изменение значимости и авторитетов
- появляются сторонники и противники
- противники выживают инноватора

Инновации. Противники инноваций

Противники инноваций боятся:

- признать, что инновационная идея не их
- того, что их фронт работ выполнят другие
- упустить признание былых заслуг
- потерять значимость, незаменимость, эксклюзивность на выполнение функций
- потерять место в компании
- не способны воплотить инновацию
- не способны пользоваться уже внедренной инновацией

Организационная структура. Тенденции

- ✓ Промышленные и финансовые организации выделяют подразделения в отдельное юридическое лицо с постфиксом «digital» или «цифра»:
- ✓ Математическое направление в составе юрлица «digital», но иногда data office выделяют тоже в отдельное юрлицо с постфиксом «DataLab»

Причины:

- Вычленить коммерческую составляющую этого юнита
- Не позволить разодрать инновационные подразделения
- Локализовать экспертизу в сервисной организации

Цели ИИ: «торт» ценности

ценность

действие

аналитика и отчетность

процесс создания достоверных данных и поддержание их в достоверном состоянии

Монетизация данных методами продвинутой математики

Данные – не самоцель — IT

Аналитика – не самоцель — МАТЕМАТИК

Знания = функция от данных — МАТЕМАТИК

Действие - основано на знаний — БИЗНЕС

Рекомендации

Не строй корпоративное КЛАДБИЩЕ данных

Представляй конечную цель – монетизацию посредством действия, направляемого знаниями, полученными аналитикой из данных

Если хочешь построить корабль, не учи людей пилить деревья, научи их мечтать о море.

Монетизация. Декомпозиция цели

- Целеполагание: потребность бизнеса
- Логическая форма: математическая формулировка, отчетность, математическая модель, BI модель
- Архитектура структуры данных
- Хранилище данных: архитектура, качество, глоссарий, метаданные, золотая карточка и тп.

Создание кладовки данных без понимания логической формы и аналитического инструмента может привести к созданию корпоративного КЛАДБИЩА данных, пусть и тщательно задокументированного

Признаки провала

- отсутствует понимание разницы между традиционной аналитикой (BI, отчетность) и продвинутой аналитикой (прогнозирование, предписание)
- опора бизнес-лидеров на интуицию и традиционную практику принятия решений, а также сопротивление изменениям
- стремление иметь технологию, а не решить задачу

Признаки провала

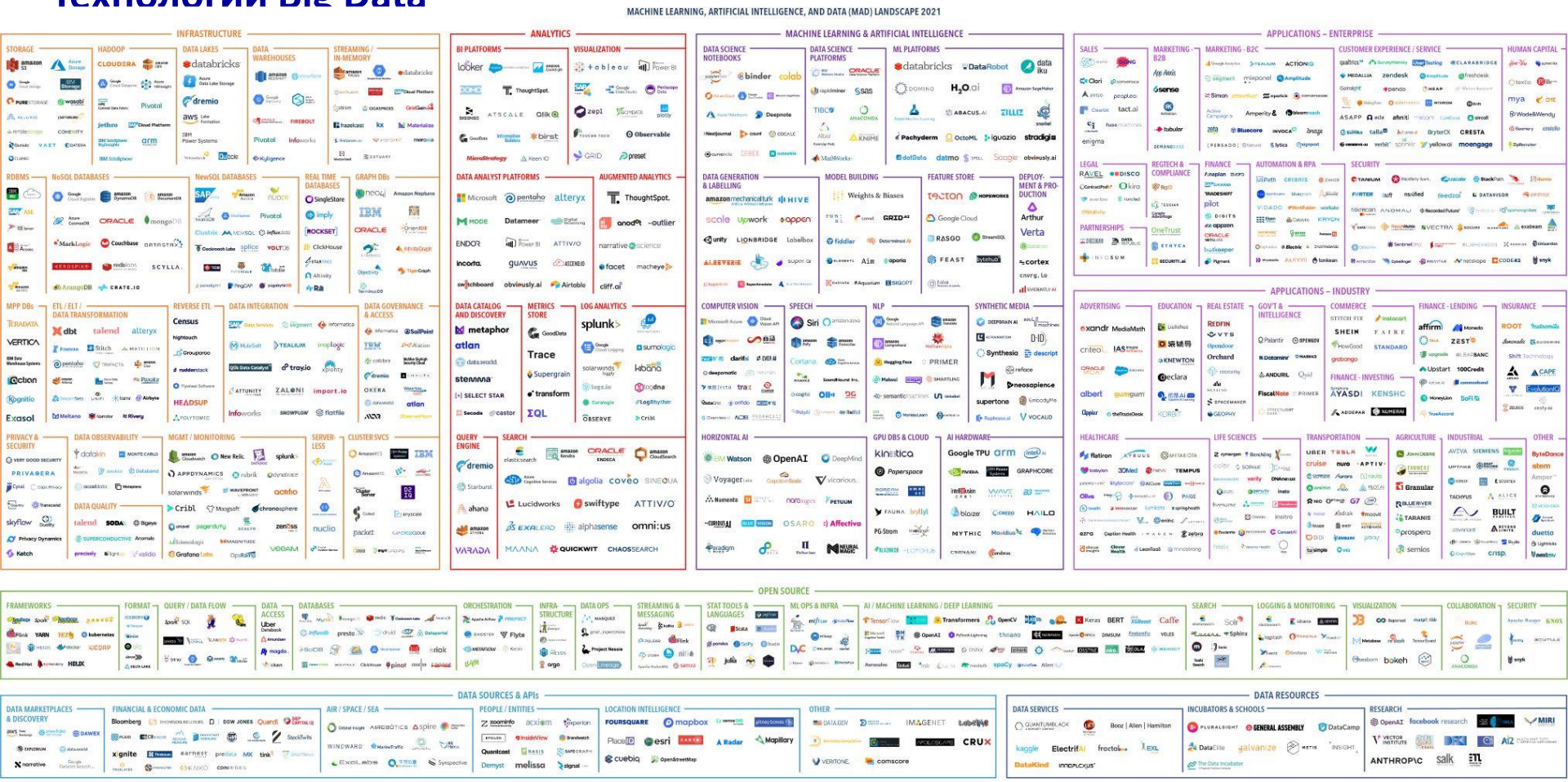
- финансовый эффект аналитической инициативы не оценен
- некачественны данные: нехватка, однобокость, проблемы интеграции, неучтенная динамика, пренебрежение неструктурированными данными
- иллюзия того, что ИИ может решить любую задачу и без применения усилий, как волшебная палочка

Признаки провала

центр аналитической компетенции изолирован от бизнеса:

- отсутствует бизнес-аналитик/специалист по монетизации, гармонизирующий взаимодействие заказчика и аналитиков данных.
- математики не знают предметной области
- невостребованность бизнесом/ заказчиками аналитических моделей

Технологии Big Data

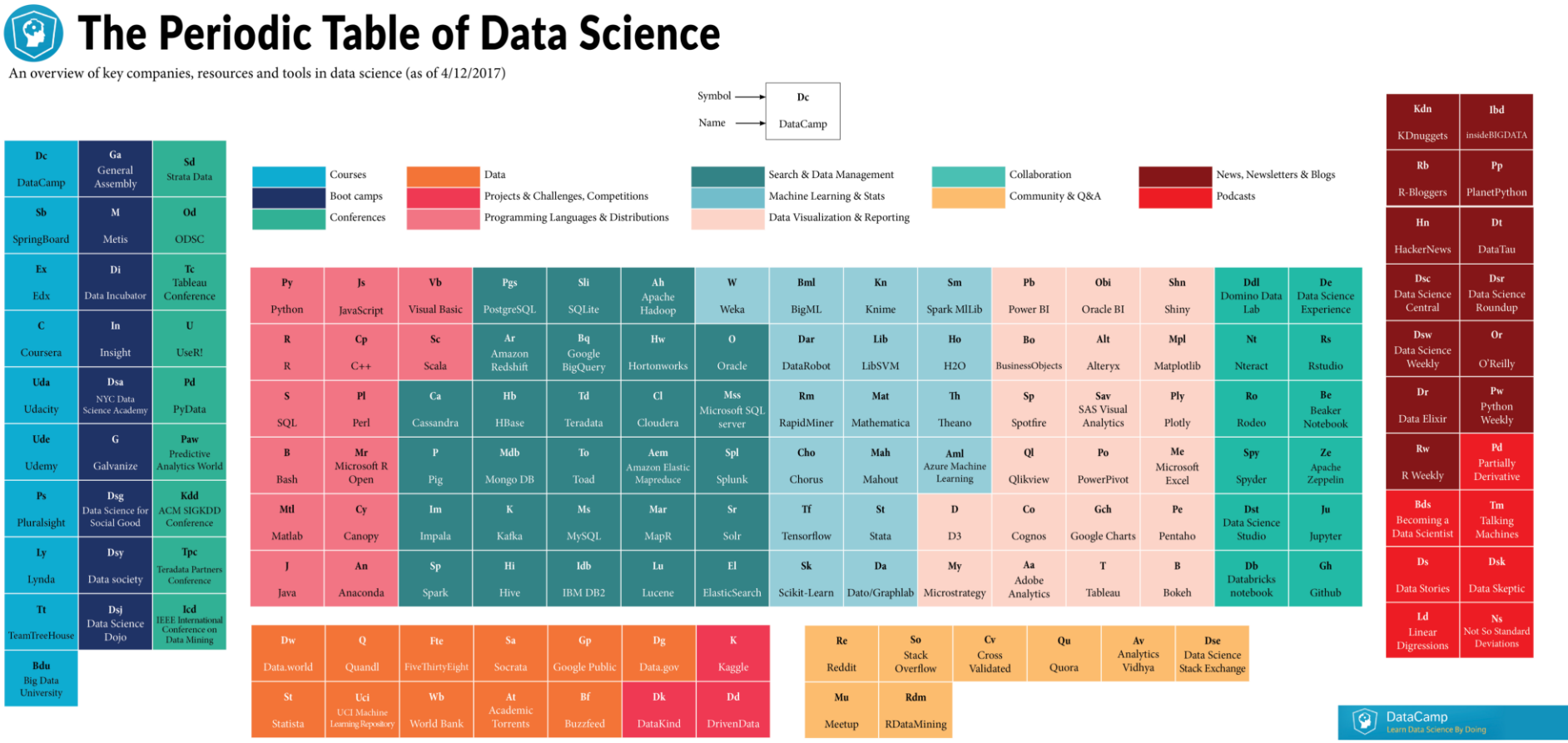


- Зоопарк технологий
- Дорогой в содержании
- Трудно найти специалистов, которые бы владели ровно Вашим стеком технологий

Рекомендации:

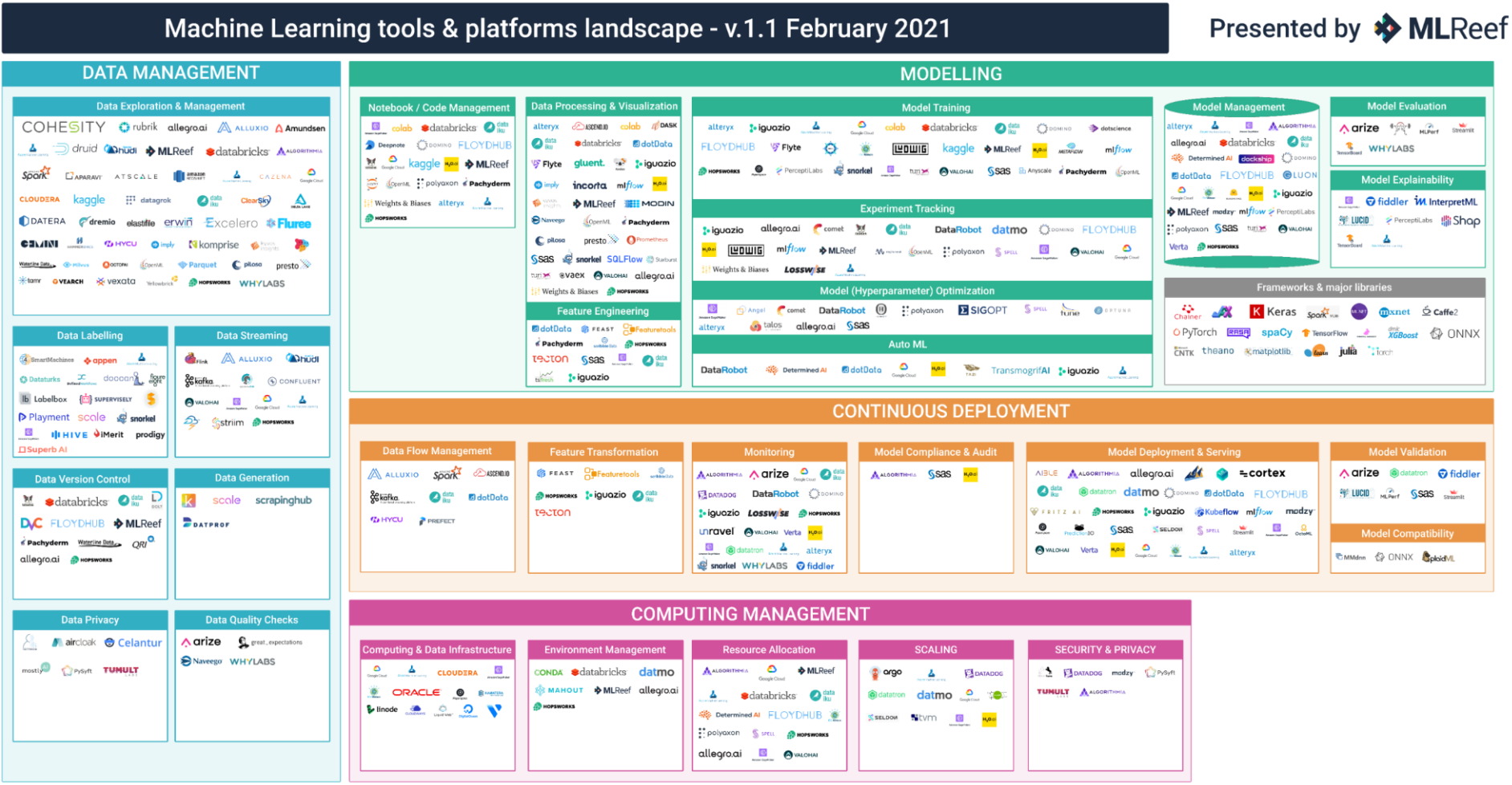
- Не гонитесь за технологиями, выбирайте наиболее распространенные
- Давайте шанс новым специалистам освоить Ваш стек. Не питайте иллюзий найти готового

Технологии анализа данных

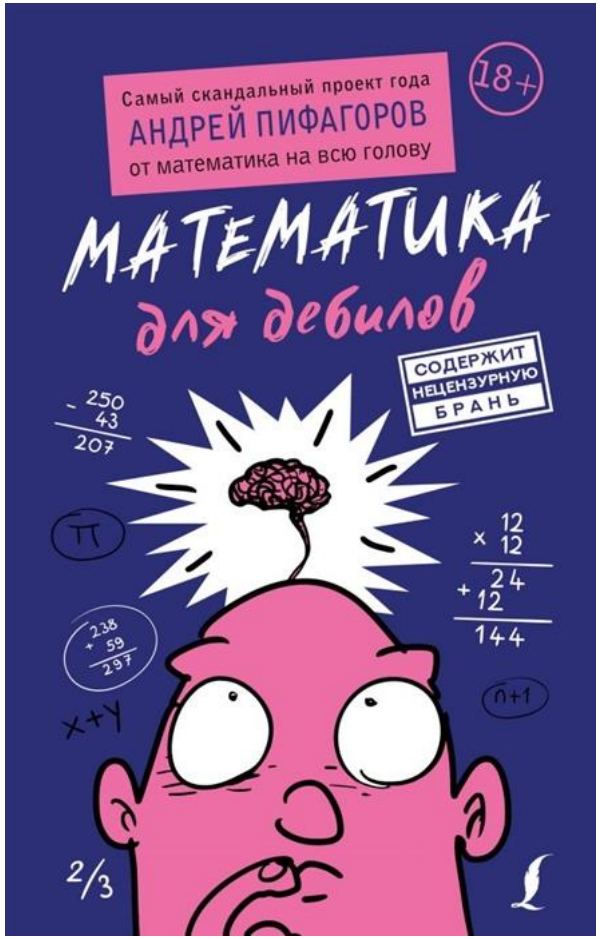


Не используйте экзотические методики анализа данных, или обучайте новых сотрудников этим методикам, не питайте иллюзий относительно поиска готового специалиста

Технологии анализа данных



О том, как стать математиком за 24 часа ☺



- Математические методы требуют профильного математического образования.
- Доступность технологий не делает новоявленного «математика за 7 дней» профессионалом

Характеристики руководителя подразделения AI

crazy scientist?

«всемогущий» менеджер-универсал?

- Математические методы (ИИ, data science, управление данными) – экспертная область, здравого смысла и общих соображений недостаточно.
- Организационная структура, регламенты, ответственность, приказы важны! Успех не должен зависеть от чудес коммуникации
- Коммуникатор???? Софт-скилы не должны сглаживать несовершенство организационной структуры, рано или поздно случится сбой
- Рекомендация: приглашайте на роль руководителя состоявшегося эксперта и прокачивайте его менеджерские скилы

Yury Sirota

PhD in Mathematics

MSc in Finance

MSc in Software engineering

Decision scientist

Contact information

+79267200790

YurySirota@ yandex.ru

YurySirotaSkype

Moscow, Russian Federation

www.linkedin.com/in/YurySirota

www.facebook.com/YuriiSirota