



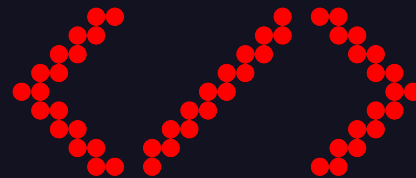
Сборка\_



**GoodLine**  
Оператор связи Кузбасса



АТВИНТА



Магия и боль ML

# Машинное обучение - это

**Машинное обучение** — класс математических методов, характерной чертой которых является не прямое решение задачи, а обучение в процессе решения множества сходных задач.

**Или** другими словами восстановление зависимостей по данным.

Только зависимости могут быть многомерными и очень сложными, а данные могут быть разнородными, неточными, неполными и даже противоречивыми.

# Основная задача машинного обучения

## Этап №1 – обучение с учителем

- **На входе:** данные – выборка прецедентов «объект -> ответ»
- **На выходе:** алгоритм, по любому объекту предсказывающий ответ

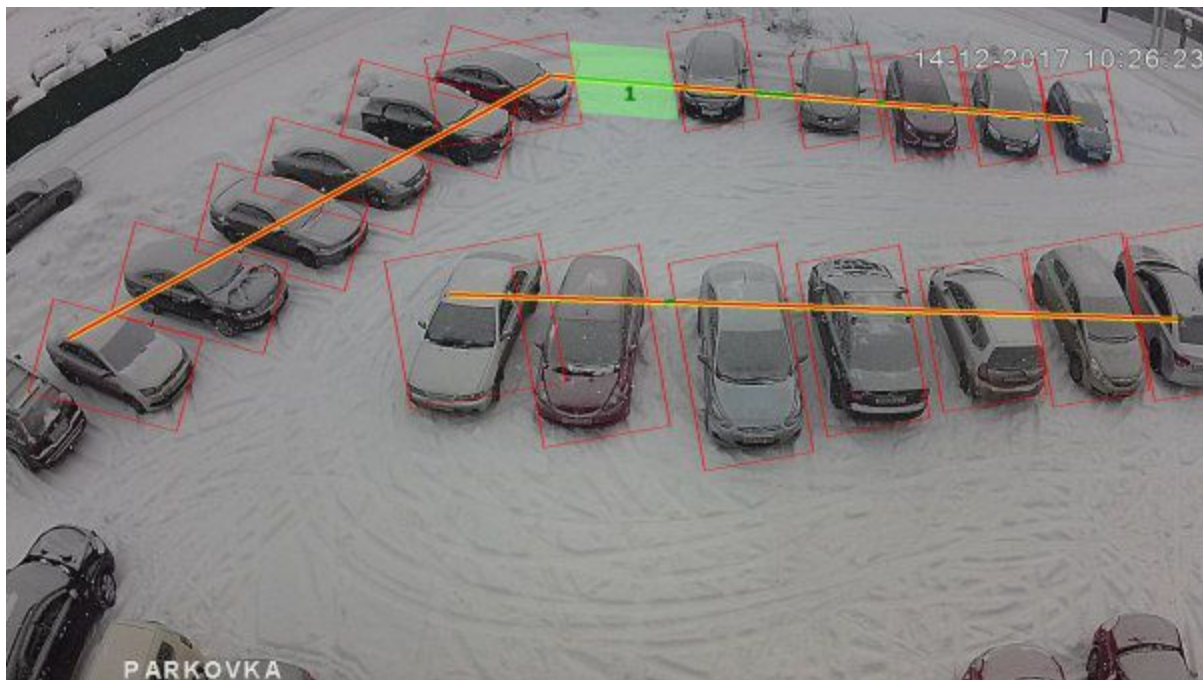
## Этап №2 – применение

- **На входе:** данные – выборка новых объектов
- **На выходе:** предсказания алгоритма на новых объектах

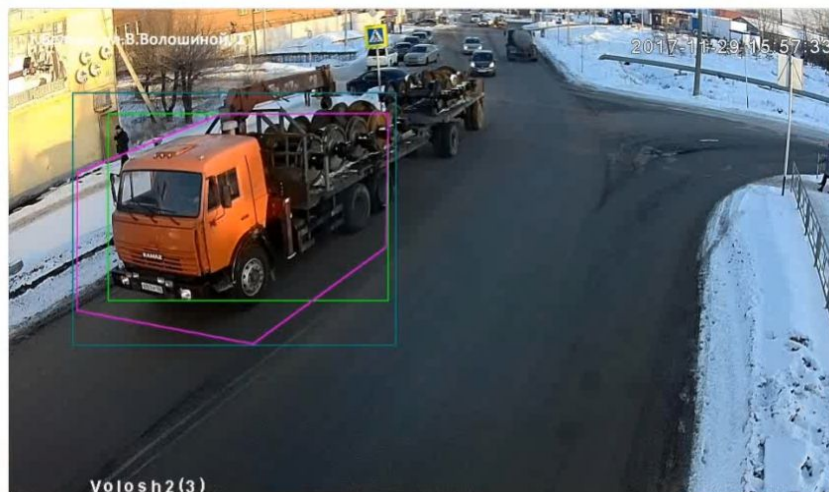
# Примеры задач машинного обучения

- Кредитный скоринг
- Предсказание оттока клиентов
- Прогнозирование объемов продаж
- Рекомендательные системы
- Управление технологическими процессами
- Медицинская диагностика
- Перевод текста
- Компьютерное зрение

# Поиск свободного места на парковке



# Детекция грузового транспорта



Предсказание оттока абонентов



# О задаче

*Отток* - абонент который в течение 3 месяцев не платит за услугу интернет

Данные:

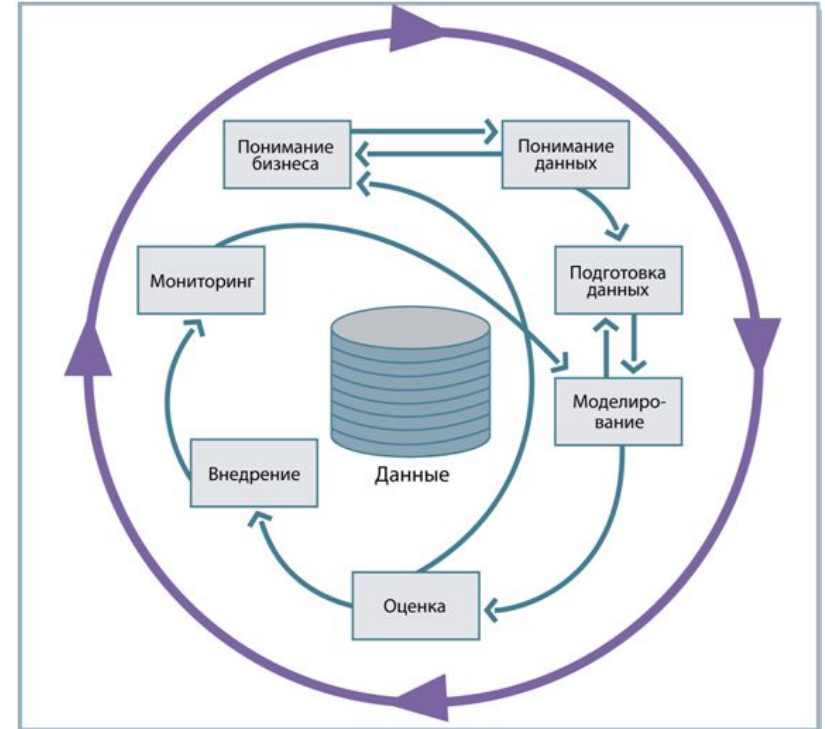
- **счета и платежи**— вся информация, связанная с финансами
- **использование услуг** — информация о трафике, продолжительность использования, ...
- **подключенные услуги** — какие, когда, как долго
- **остальное**— география, коммуникации с контакт-центром, ...

# Цикл работы

Этапы:

1. Понимание бизнеса(формулирование цели и задач)
2. Работа с данными
3. Моделирование
4. Внедрение
5. Мониторинг

Основной цикл работы - CRISP-DM.

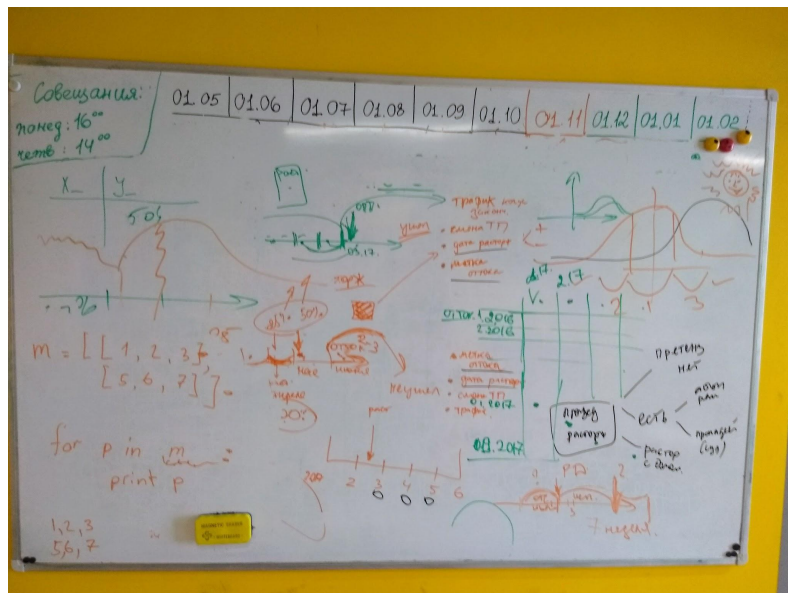


# Заведите глоссарий

Зафиксируйте ключевые понятия и решаемую задачу.

Дополните рисунками если необходимо.

Пусть каждый прочитает, перескажет.



# Этапы работы с данными

Этапы работы с данными:

1. **Выборка данных:** что, откуда, есть ли исторические данные, соотнесение с объектами
2. **Очистка:** устраняем ошибки, заполняем пропуски
3. **Преобразование:** сгруппировать помесячно

# Работа с данными это 80% времени

С данными обязательно что-то не так!

1. Ошибки ввода
2. Пропуски
3. Аномалии
4. Дубли



# Понимание данных

Узнайте у экспертов максимально о предметной области и о решаемой задаче.

Задавайте вопросы:

- Как задача решается сейчас?
- Какие факторы влияют?
- Какие значения может принимать признак?
- Как меняется динамика от месяца к месяцу?
- Какие были изменения в процессах и как это изменилось на данных?

# Гипотезы

**Гипотеза** – это предположение, о зависимости целевой переменной и какого-то фактора.

Гипотеза должна быть проверяемой и измеримой.

Сформулировали -> Провели эксперимент -> Оценили  
-> Отвергли/приняли



# Больше гипотез!

Придумывайте как можно больше идей(гипотез) и проверяйте их.

80% гипотез будут отвергнуты или не принесут практического результата!

Записывайте гипотезы в реестр.

Тюнинг моделей - даст прирост в 3-10%.

**Качественное улучшение можно получить за счет “новых” признаков.**



# Моделирование

Решайте задачу итеративно.

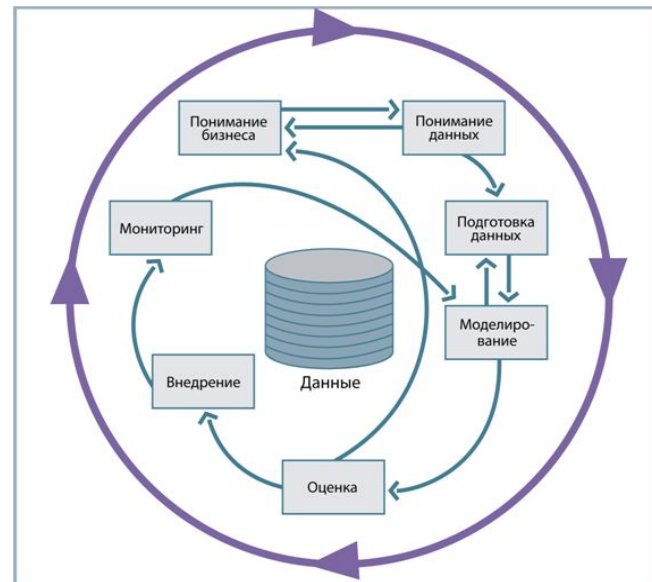
1. Сделайте прототип
2. Проверьте его
3. Улучшите или начните заново

Интерпретируйте модель.

Начинайте с более простых моделей.

Каждый полученный результат интерпретируйте:

- Что он значит?
- Какая практическая польза?
- Помогает ли это достигнуть цели?



# Технологии которые используем



# Заключение

# Достижения

1. Беспилотный автомобиль DARPA (2005)
2. Apple Siri, голосовой помощник (2011)
3. WATSON выиграл в “Jeopardy” (2014)
4. ImageNet, аннотирование изображений с точностью выше чем у людей (2014)
5. DeepMind. OpenAI: обучение компьютера играм Atari (2016)
6. Microsoft, распознавание голоса компьютером превысило человеческий уровень (2016)
7. Google DeepMind обыграл чемпиона по игре в го (2016)

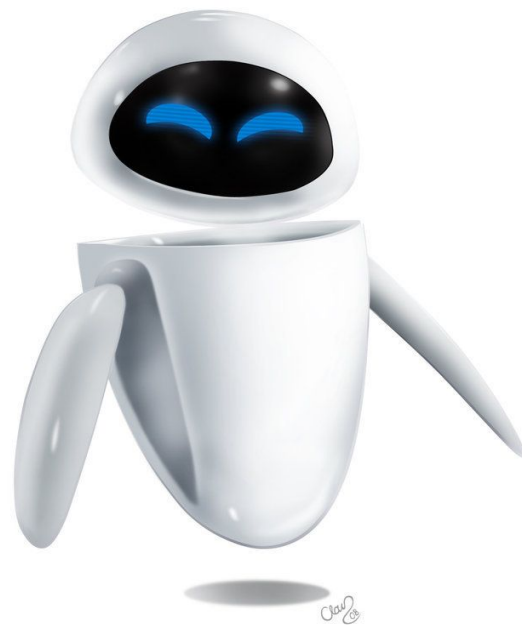
## Технологии, меняющие жизни людей

1. Накопление больших объемов данных
2. Удешевление и миниатюризация вычислительных ресурсов, датчиков, аккумуляторов
3. Удешевление робототехники
4. Успехи в развитии методов машинного обучения

За счет этих факторов происходит стремительная автоматизации труда.

## Тренды

1. Распространение умных вещей
2. Беспилотный транспорт
3. Улучшение качества сервисов и услуг за счет персонализации
4. Интеллектуализация машинного труда: профессии где решения принимаются на основании типовых факторов будут заменяться машинами



## С чего начать

- «[Введение в машинное обучение](#)» на Coursera
- Специализация «[Машинное обучение и анализ данных](#)» на Coursera
- Сообщество [OpenDataScience](#) в Slack

Спасибо за внимание

Доронин Евгений, telegram: @ded42