Машинное обучение

Проскуряков Родион Александрович, 2019

http://tiny.cc/204tez

Что это такое?

Машинное обучение - это область науки, занимающаяся исследованием алгоритмов и статистических моделей, которые компьютеры используют для выполнения конкретных задач, без четких инструкций. Алгоритмы машинного обучения опираются на данные, в которых они самостоятельно ищут закономерности.

Зачем это все нужно?

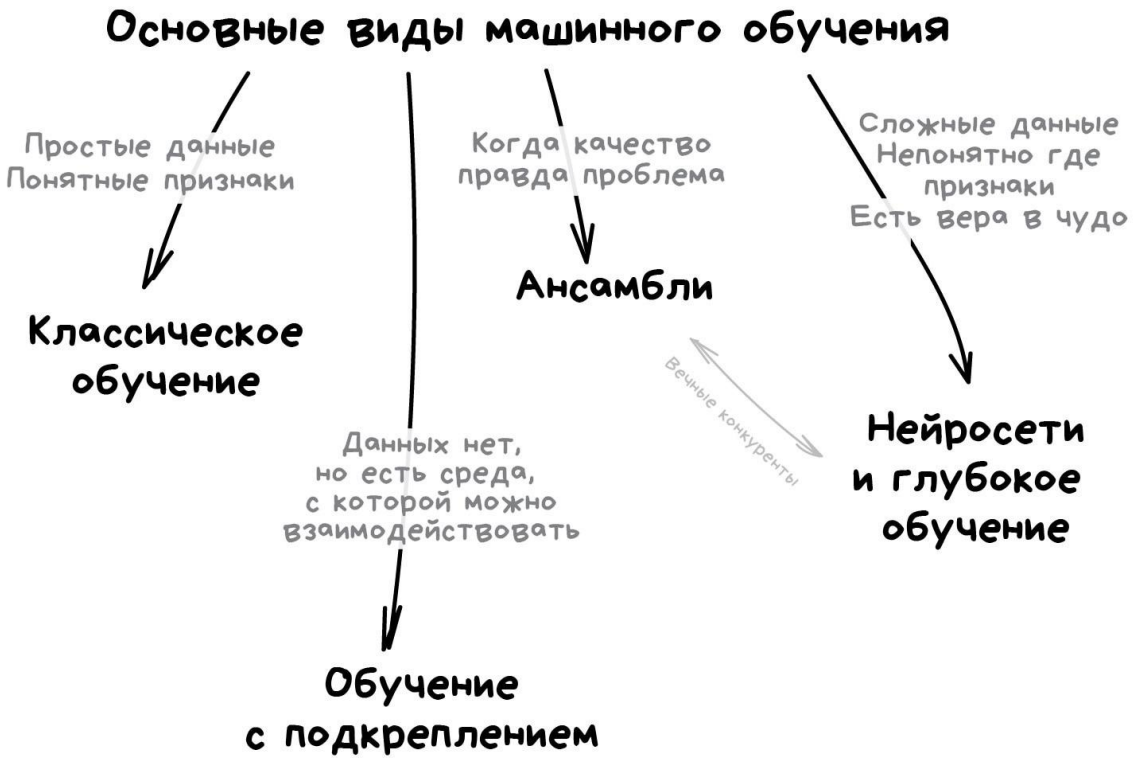
Хотим имея какие-то данные, научить машину делать предсказания, т. е. чтобы машина нашла **зависимость.**

Знаем, как что в часе 60 минут: М = 60Н

А как зависит цена квартиры от ее площади, числа комнат, расположения относительно магазинов и т. д.?

А что про искусственный интеллект?

Основные виды машинного обучения



Как машины ведут себя при пожаре

Классическое программирование

«Я просчитал все варианты событий и ты сейчас должен связать верёвку из хлебного мякиша»

Машинное обучение

«По статистике люди гибнут в 6% пожаров, поэтому рекомендую вам умереть прямо сейчас»

Обучение с подкреплением

«Да просто беги от огня АААААААААА!!!!
»

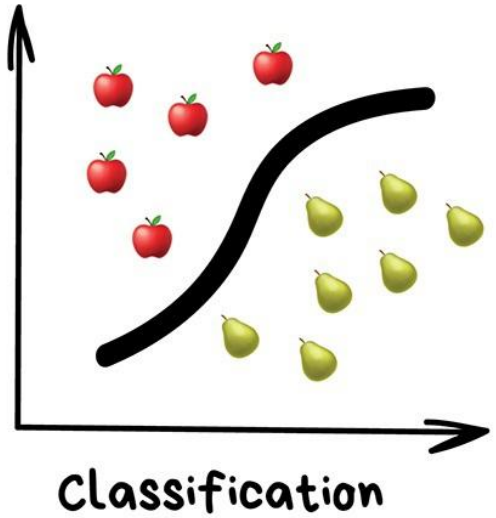
Классическое обучение

Классификация

«Разделяет объекты по заранее известному признаку. Носки по цветам, документы по языкам, музыку по жанрам»

Сегодня используют для:

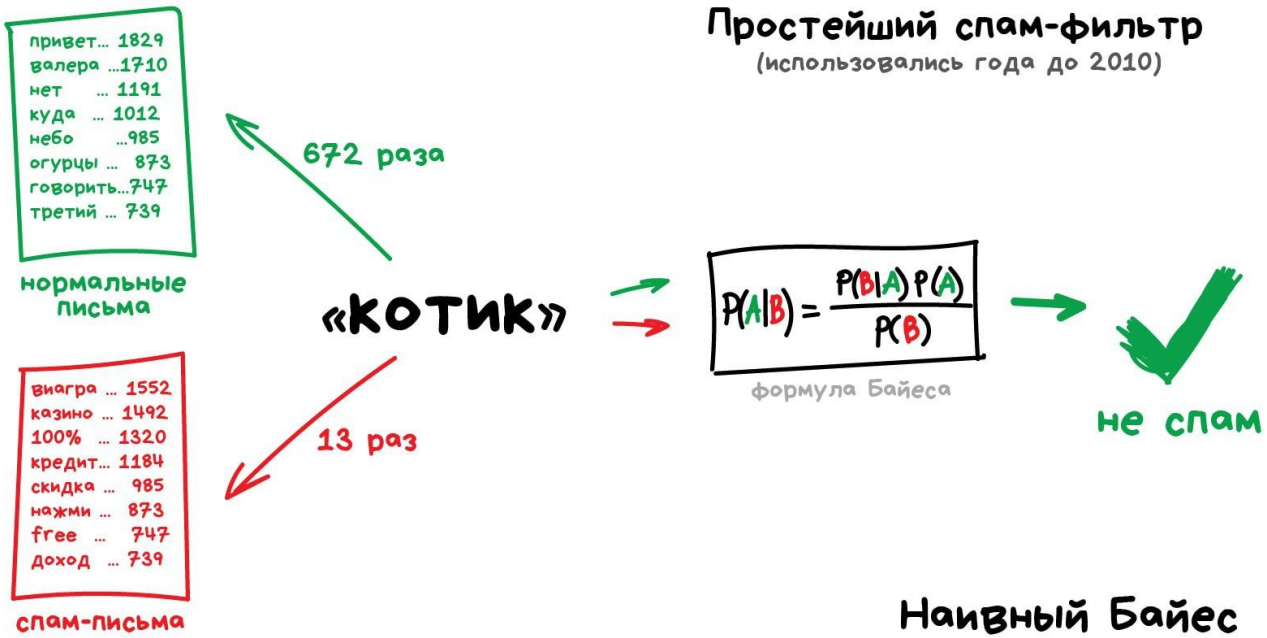
- Спам-фильтры
- Определение языка
- Поиск похожих документов
- Анализ тональности
- Распознавание рукописных букв и цифр
- Определение подозрительных транзакций



Деревья решений

Дер

Наивный Байес

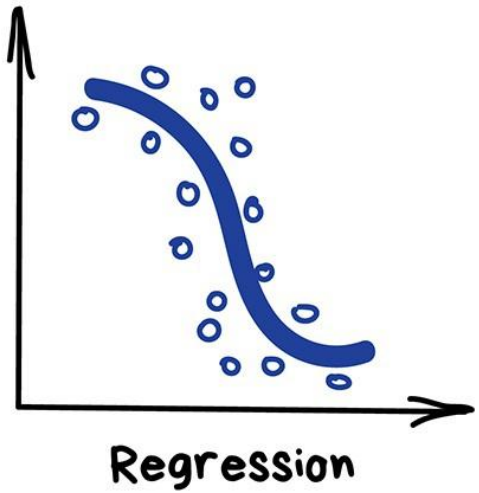


Задача регрессии

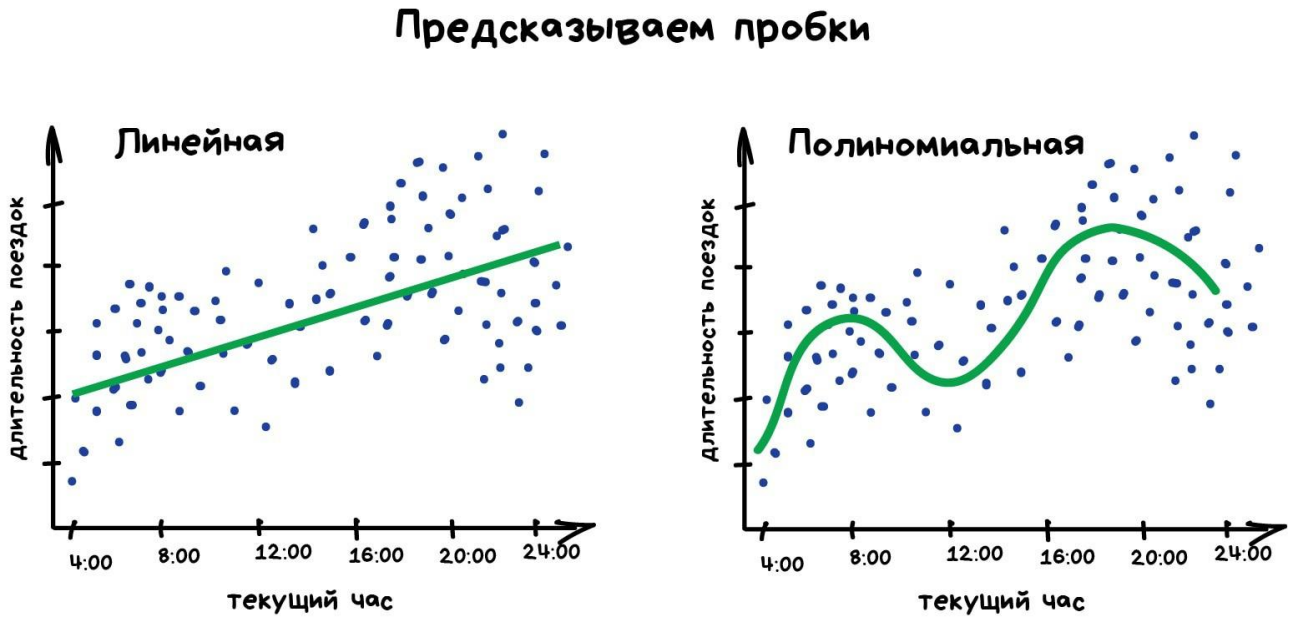
«Нарисуй линию вдоль моих точек. Да, это машинное обучение»

Сегодня используют для:

- Прогноз стоимости ценных бумаг
- Анализ спроса, объема продаж
- Медицинские диагнозы
- Любые зависимости числа от времени



Пример задачи регрессии

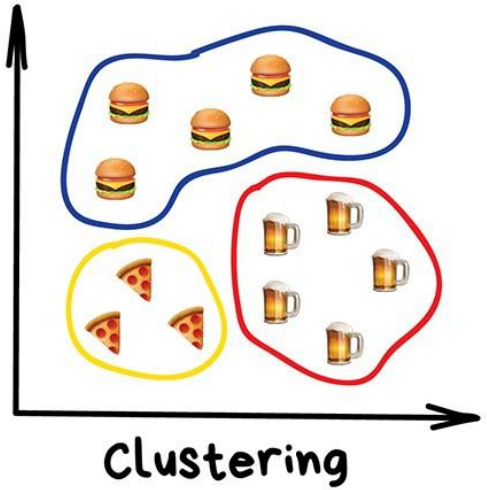


Задача кластеризации

«Разделяет объекты по неизвестному признаку. Машина сама решает как лучше»

Сегодня используют для:

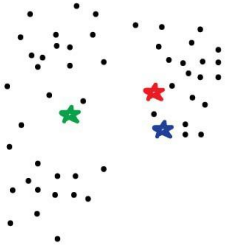
- Сегментация рынка (типов покупателей, лояльности)
- Объединение близких точек на карте
- Сжатие изображений
- Анализ и разметка новых данных
- Детекторы аномального поведения



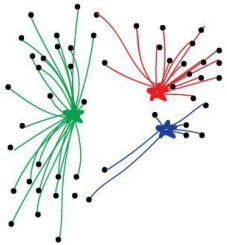
Кластеризация

Ставим три ларька с шаурмой оптимальным образом

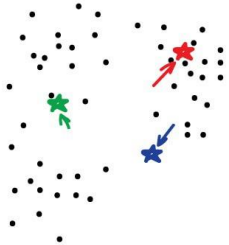
(иллюстрируя метод K-средних)



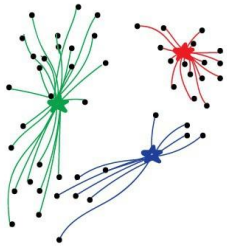
1. Ставим ларьки с шаурмой в случайных местах



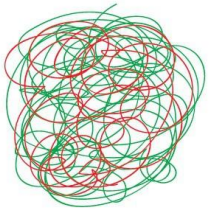
2. Смотрим в какой кому ближе идти



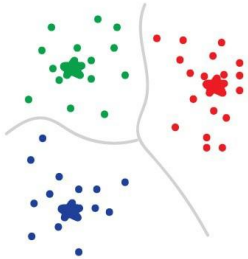
3. Двигаем ларьки ближе к центрам их популярности



4. Снова смотрим и двигаем



5. Повторяем много раз



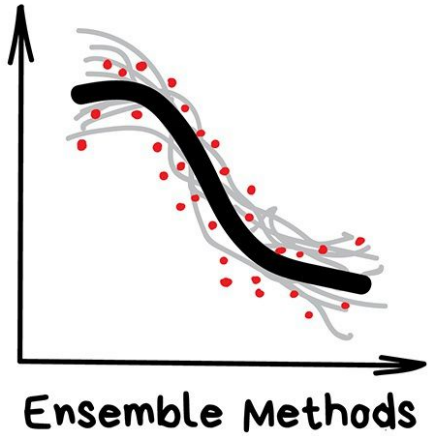
6. Готово, вы великолепны!

Ансамблирование алгоритмов

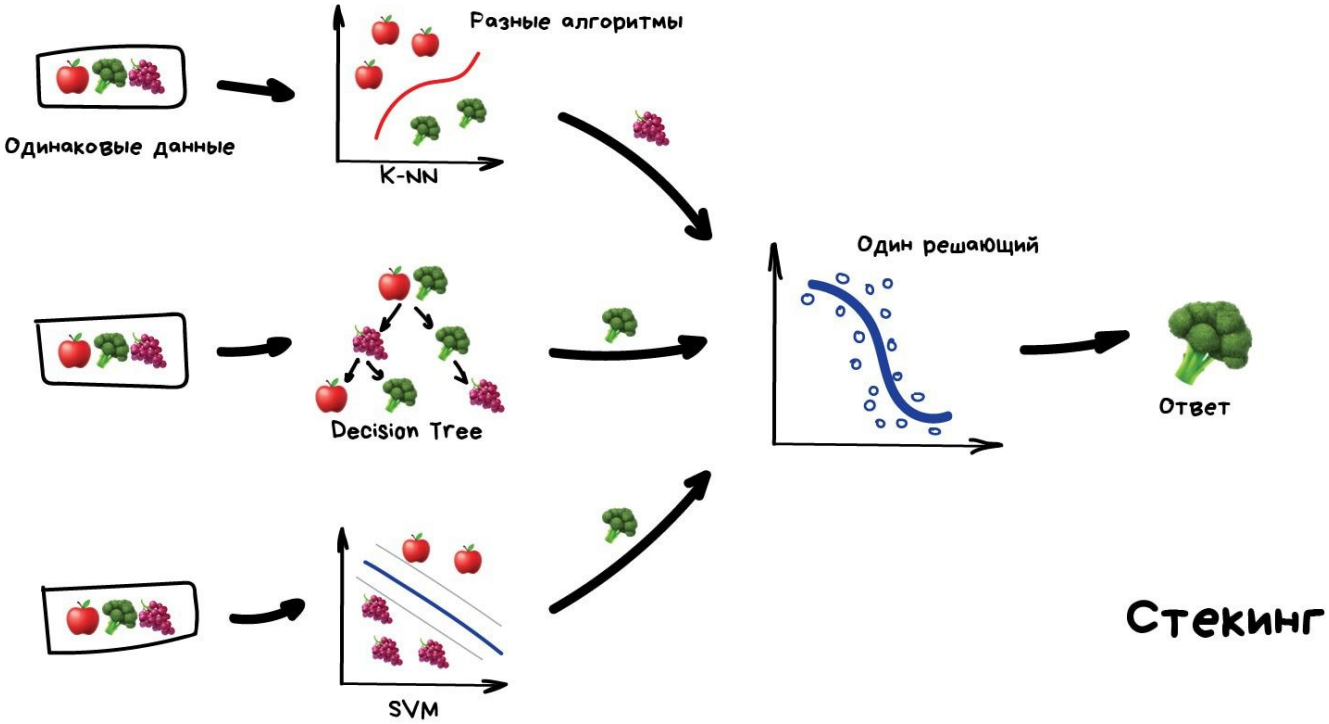
«Куча глупых деревьев учится исправлять ошибки друг друга»

Сегодня используют для:

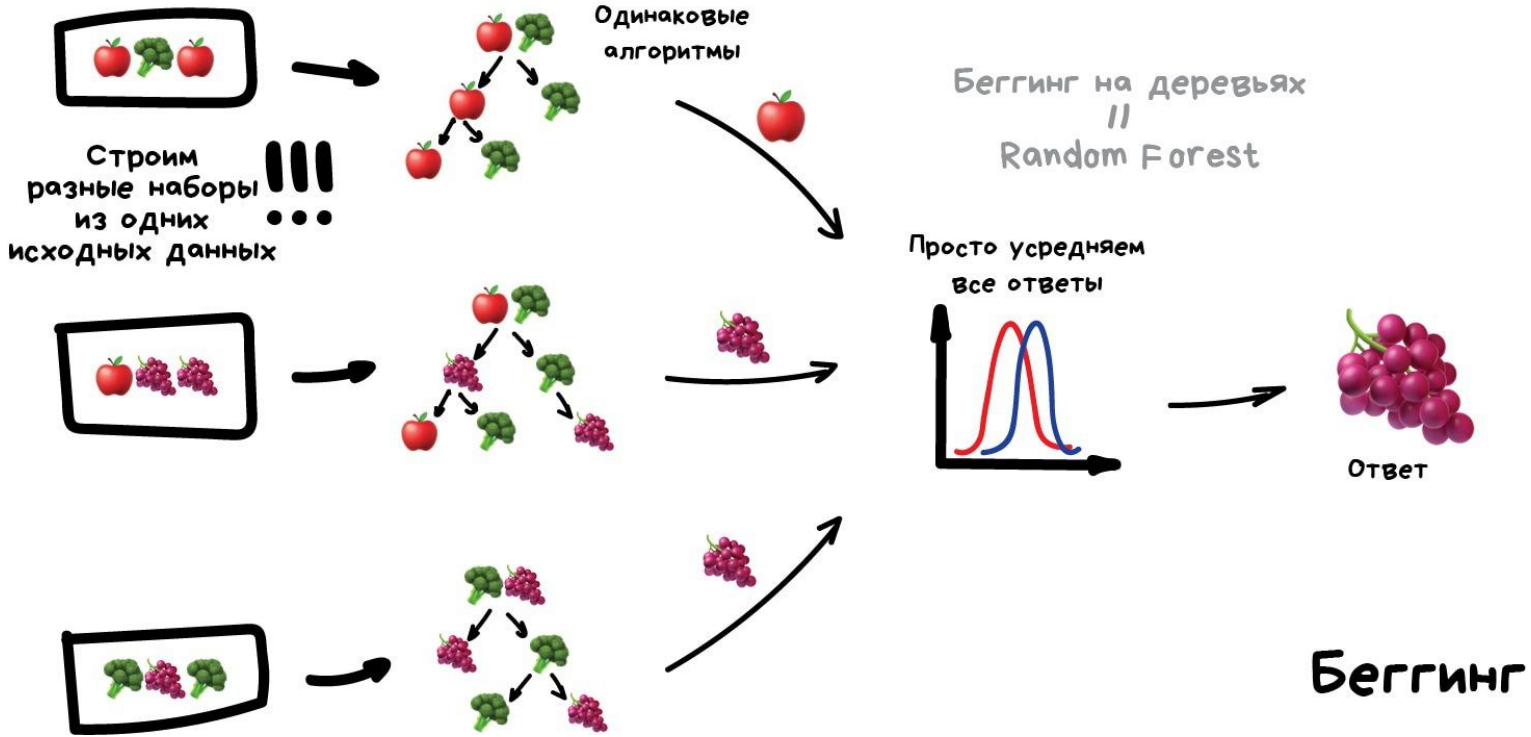
- Всего, где подходят классические алгоритмы (но работают точнее)
- Поисковые системы (★)
- Компьютерное зрение
- Распознавание объектов



Ансамблирование: стекинг

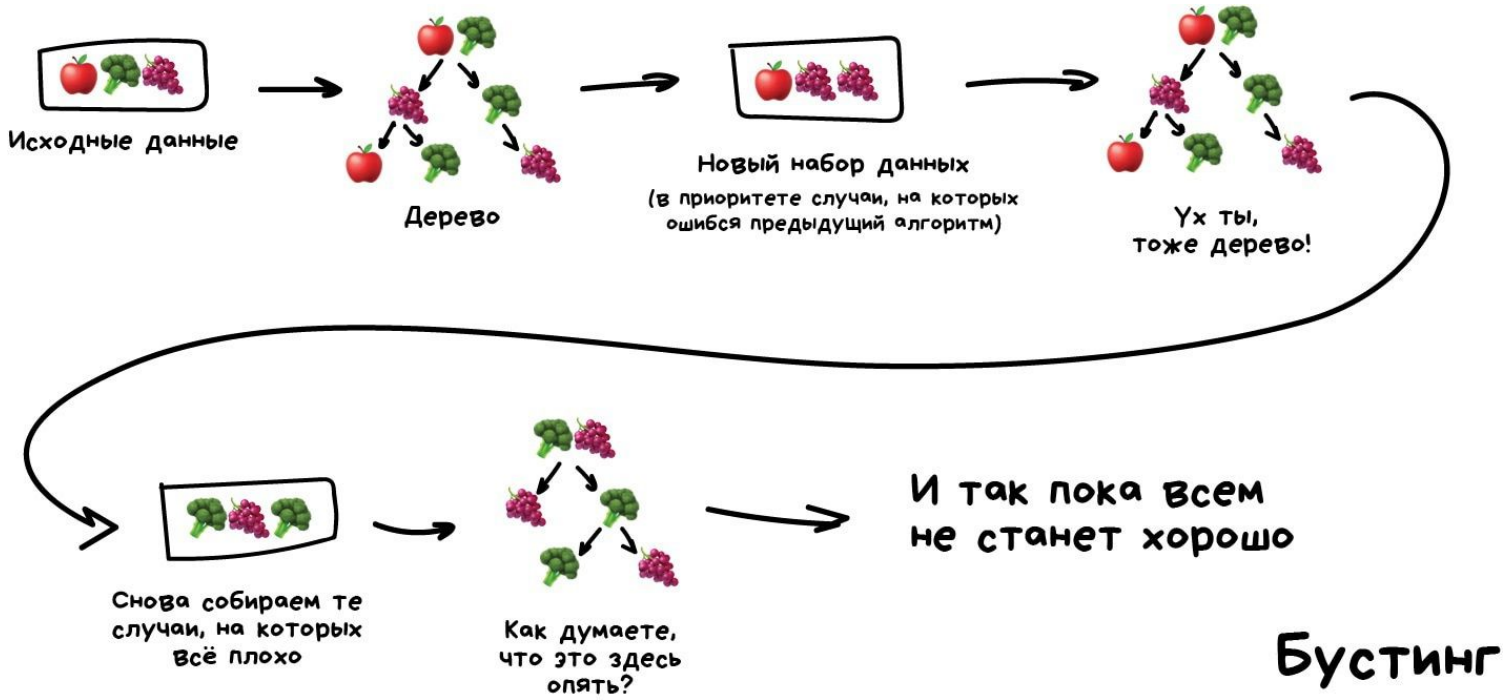


Ансамблирование: беггинг



Пример применения беггинга

Ансамблирование: бустинг



Что нужно знать, для того, чтобы заниматься ML?

1. Python

   Курс на stepik (если вы раньше не программировали)

   Курс на stepik (если уже знакомы с другим языком)

2. Немного математики
3. Классические алгоритмы (этим мы и будем заниматься вначале)
4. Нейронные сети? (это будет попозже)