

Машинное обучение на практике

Евгений Соколов

esokolov@hse.ru

Вороново, 27 января 2018

Успехи в области специализированного ИИ

Нейронная сеть, играющая в Го

- **Март 2016** — победа над мировым чемпионом
- Нейронная сеть обучалась, играя сама с собой для увеличения объёмов входных данных (принцип обучения с подкреплением, reinforcement learning)



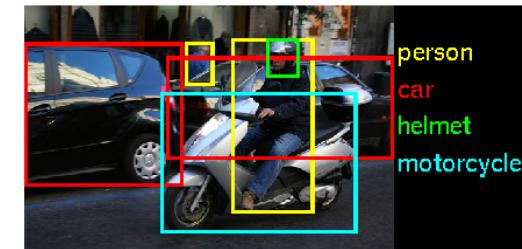
Вопросно-ответная система, построенная методами машинного обучения

- Извлекает ответы из больших объёмов структурированных (WordNet) и неструктурных (Wikipedia) данных
- **Февраль 2011** — победа в игре «Jeopardy!»



ImageNet — задача распознавания объектов на изображении

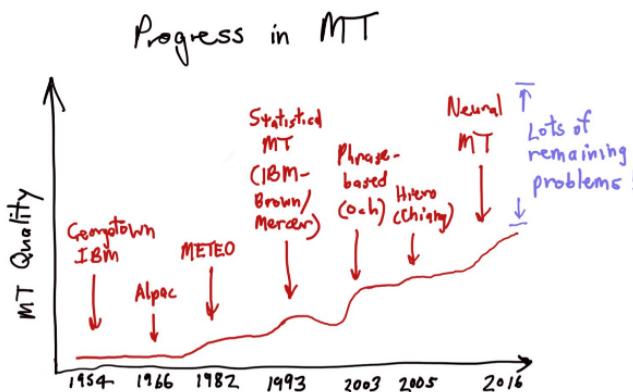
- Решается с помощью нейронных сетей с точностью, превышающей точность работы человека
- Сделан рывок в области методов переноса знаний между моделями
- Применения в беспилотных автомобилях



Успехи в области специализированного ИИ

Нейронный машинный перевод

- End-to-end модели
- Обошли по качеству модели на основе правил и статистические модели, которые развивались с середины 20-го века



Рекомендательные системы

- Применяются повсеместно для рекомендаций товаров, медийного контента, новостей, людей в социальных сетях, ресторанов и достопримечательностей и т.д.
- Применения: upsell в финансовом секторе, коммуникации с должниками
- Используют информацию об активности всех пользователей сервиса, а также анализируют содержимое рекомендуемого контента



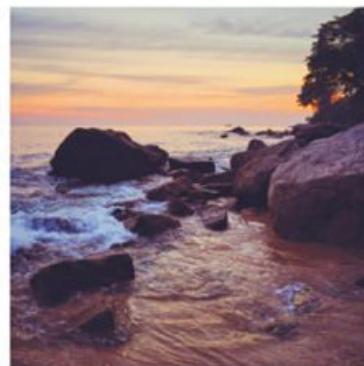
Генеративные модели и анализ текста

- Порождение изображений и музыки
- Генерация текста по изображению и изображения по тексту
- Перенос стиля
- Чат-боты и автоматизация службы поддержки
- Автоматизация поиска сайтов МФО в интернете



Машинное обучение в
искусстве и дизайне

Машинное обучение в искусстве и дизайне



Машинное обучение в искусстве и дизайне

- Генерация музыки и видео
- Генерация текстов и поэзии

VIOLA:

Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:

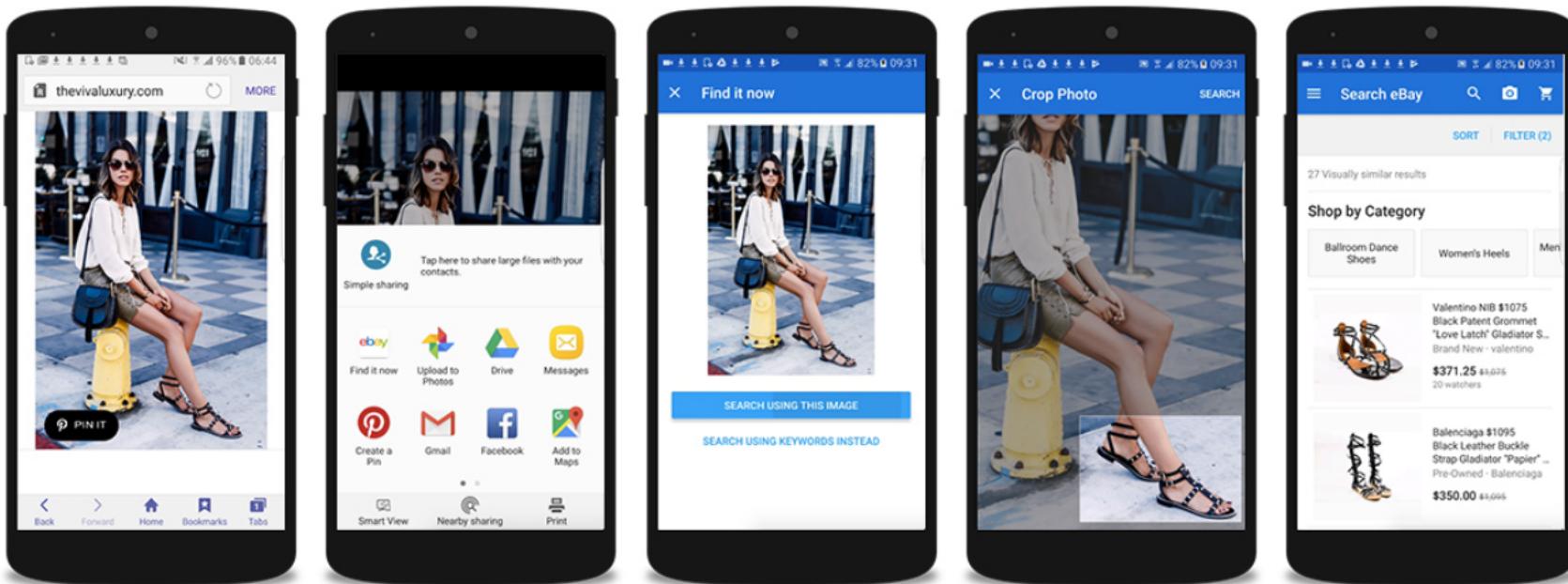
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.





Машинное обучение в искусстве и дизайне

- Генерация новых стилей одежды и классификация стиля
- Поиск товаров по фотографии



ebay

Машинное обучение в искусстве и дизайне

- Новые инструменты для творчества
- Новые возможности в fashion
- Искусственный интеллект не может придумывать принципиально новое в искусстве
- Задача автора — придумывать новые концепции, а не заниматься рутинной работой

Рекомендательные системы,
маркетинг и медиа

Интересные предложения

РЕКОМЕНДУЕМ



Кулеры и системы охлаждения д...
[Deepcool NEPTWIN V2](#)

от 2 548 ₽

ХИТ ПРОДАЖ



Защитные пленки и стекла для т...
[Защитное стекло Xiaomi](#)

от 149 ₽

ТОВАР НЕДЕЛИ · ДО -20%



Электрообогреватели и тепловы...
[Ballu BOH/CL-07](#)

от 1 630 ₽

СКИДКИ ДО -10%



Цифровые бытовые метеостанции
[Oregon Scientific LW301](#)

от 8 990 ₽

РЕКОМЕНДУЕМ



Жесткие диски, SSD и сетевые н...
[Samsung MZ-7KE512BW](#)

Территория Grohe



ХИТ ПРОДАЖ



Дрели, шуруповерты, гайковерты
[Интерскол DA-12EP-01 ...](#)

РЕКОМЕНДУЕМ



Жесткие диски, SSD и сетевые н...
[Samsung MZ-7KE1T0BW](#)

Похожие товары



Canon EOS 200D Kit

от 32 390 ₽

1 отзыв 109 предложений

Любительская зеркальная
фотокамера

Байонет Canon EF/EF-S

Объектив в комплекте, модель
уточняйте у продавца

Матрица 25.8 МП (APS-C)

Цвет:

Цены 109



4.0

Canon EOS 750D Kit

от 32 650 ₽

5 отзывов 133 предложения

Любительская зеркальная
фотокамера

Байонет Canon EF/EF-S

Объектив в комплекте, модель
уточняйте у продавца

Матрица 24.7 МП (APS-C)

Цвет:

Цены 133



Canon EOS M10 Kit

от 21 250 ₽

4 отзыва 89 предложений

Фотокамера с поддержкой
сменных объективов

Байонет Canon EF-M

Объектив в комплекте, модель
уточняйте у продавца

Матрица 18.5 МП (APS-C)

Цвет:

Цены 89



до -30%



Canon EOS M6 Kit

от 38 300 ₽

до 84 940 ₽

2 отзыва 105 предложений

Фотокамера с поддержкой
сменных объективов

Байонет Canon EF-M

Объектив в комплекте, модель
уточняйте у продавца

Матрица 25.8 МП (APS-C)

Цвет:

Цены 105



С этим товаром часто покупают



4.0

HIPER MP15000

Универсальные внешние аккумулято...

от 1 453 ₽

30 отзывов 33 предложения

Цвет: ●

[Цены 33](#)[Цены 66](#)[Цены 16](#)[Цены 4](#)

2.5

Apple EarPods (Lightning)

Наушники и Bluetooth-гарнитуры

от 1 590 ₽

6 отзывов 66 предложений

Цвет: ○



Сетевая зарядка Apple

Зарядные устройства и адаптеры

от 642 ₽

16 предложений

Производитель: Apple

Тип: сетевая зарядка

Разъем подключения: for Apple
(Lightning)



Smart cover Apple

Чехлы для планшетов

от 3 080 ₽

4 предложения

Производитель: Apple

Совместимость: Apple

Тип: smart cover

Функция подставки: Да

Цвет: ● ●

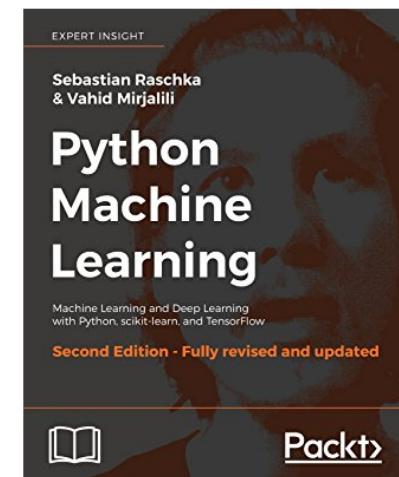
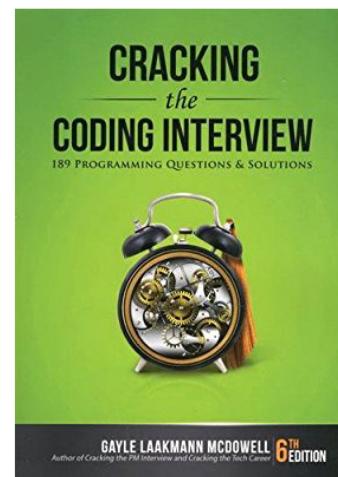
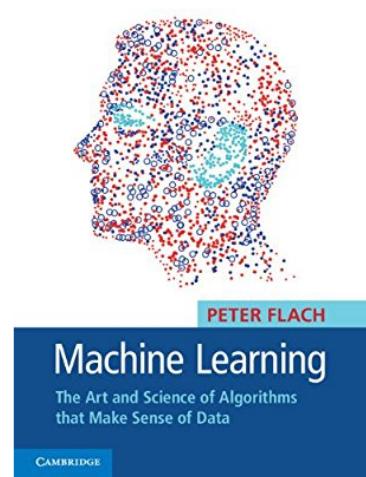
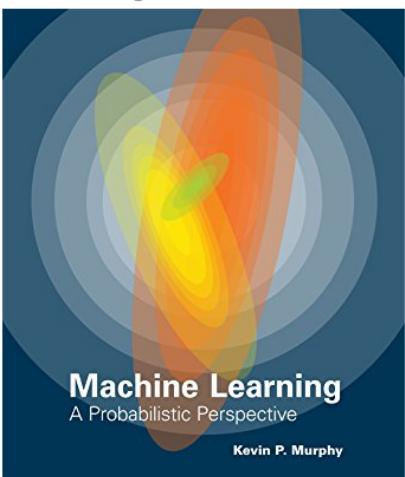
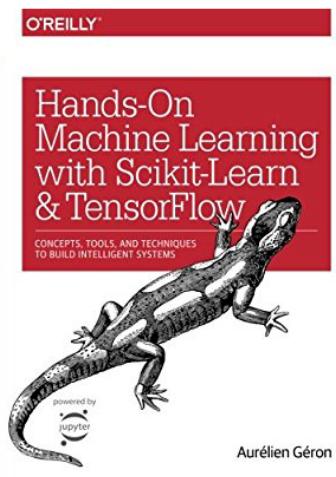
Amazon

Try Amazon Prime today and get unlimited fast, FREE shipping [See more](#)



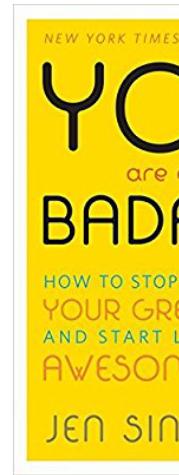
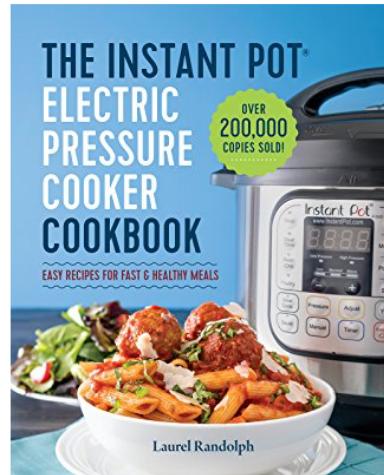
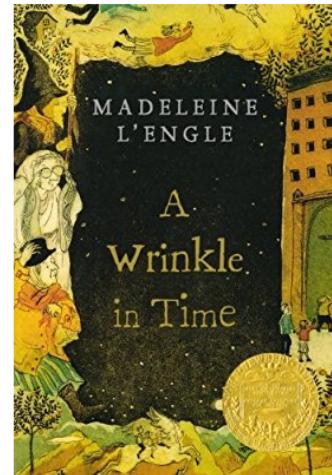
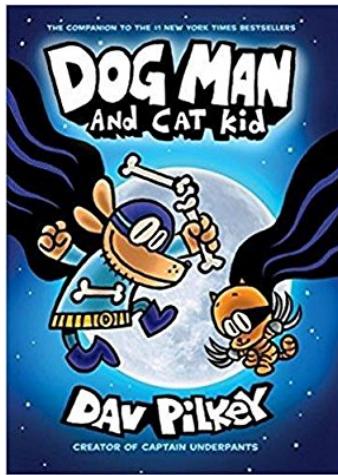
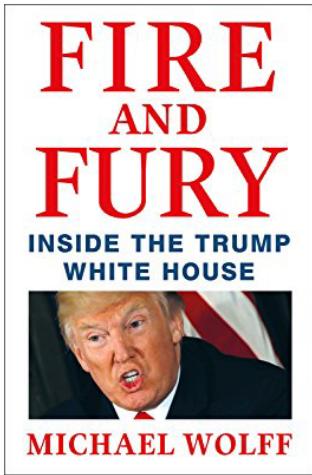
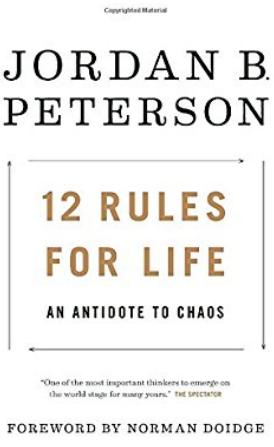
Amazon

Recommendations for you in Books



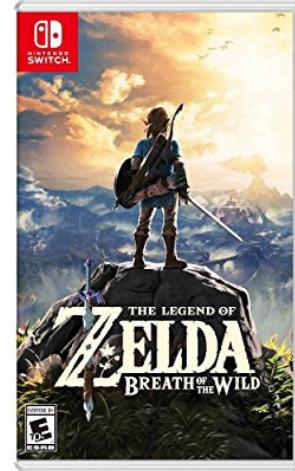
Amazon

Books best sellers [See more](#)



Amazon

Most wished for in Video Games [See more](#)



Amazon

Home & Kitchen best sellers [See more](#)



Amazon

- Большое разнообразие полок рекомендаций: персональные/популярные товары в интересных категориях, популярные категории, похожие на недавно просмотренные, аксессуары к недавно купленным и т.д.
- Рекомендации дают 35% от всех покупок

Netflix

House of Cards

★★★★★ 2013 TV-MA 1 Season HD 51

Sharks gliding ominously beneath the surface of the water? They're a lot less menacing than this Congressman.

This winner of three Emmys, including Outstanding Directing for David Fincher, stars Kevin Spacey and Robin Wright.

NETFLIX

Because you watched Orange Is the New Black

Because you watched Red Lights

Netflix

NETFLIX Browse ▾ Kids DVD NEW EPISODES

Because you watched New Girl

TV Mysteries

Irreverent TV Shows

Netflix

- Персональные рекомендации по категориям и фильмам/сериалам с интерпретациями (вам понравилось А, поэтому посмотрите В)
- 75% просмотров — благодаря рекомендациям

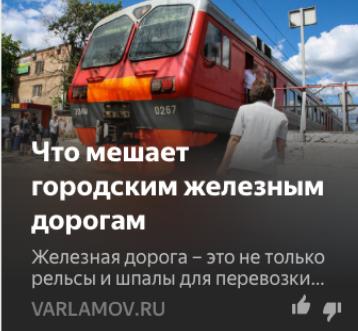
Другие примеры

- Spotify: формирование персональных плейлистов на основе поиска похожих пользователей
- Youtube: рекомендации приносят 60% от всех просмотров

Рекомендательные системы

- Персональные предложения для каждого пользователя на основе большого числа факторов
- Учёт статистики по всем пользователям сервиса
- Новый взгляд на маркетинг
- Работа маркетолога — правильное использование рекомендательных механизмов, выбор каналов для взаимодействия ИИ и пользователей

Рекомендации контента



Что мешает городским железным дорогам
Железная дорога – это не только рельсы и шпалы для перевозки...
[VARLAMOV.RU](#)

13 сюрпризов неайтишной компании
Отработав много лет в компаниях, занимающихся разработкой программного обеспечения на заказ, невольно начинаешь задумываться — а как всё-таки выглядит вся эта суэта с...
[ХАБРАХАБР](#)



Почему настоящие лидеры должны уметь думать медленно
Есть известная (но, возможно, апокрифическая) история о том, как теория относительности пришла в голову Альберту Эйнштейну, когда он ехал на велосипеде. Говорят, что Уоррен Баффетт читает не меньше шести часо...
[HBR-RUSSIA.RU](#)



Математики обыграли букмекеров
Букмекеров можно обыгрывать, если учесть, что они иногда выставляют не самые выгодные для себя коэффициенты, чтобы снизить риски. Международная груп...
[N+1](#)

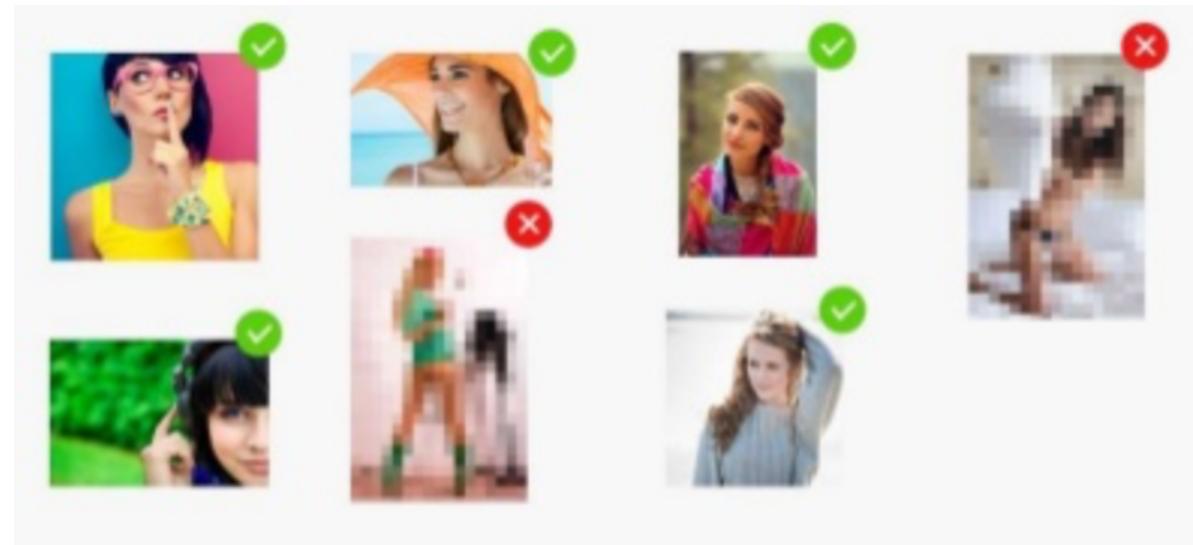
Рекомендации контента

- Медийный бум приводит к взрывному росту объёмов информации в сети
- Рекомендательные системы помогают ориентироваться
- Для авторов — поиск целевой аудитории
- Пионеры в Китае — Toutiao (более 100 миллионов активных пользователей) и другие платформы

Бизнес и автоматизация

Автоматическая модерация

- Сервис онлайн-знакомств
- Фото в профиле не должно содержать эротику и не должно быть фотографией знаменитости
- Точность 88%
- Экономия порядка \$1,000,000 в год



Рекомендательные системы

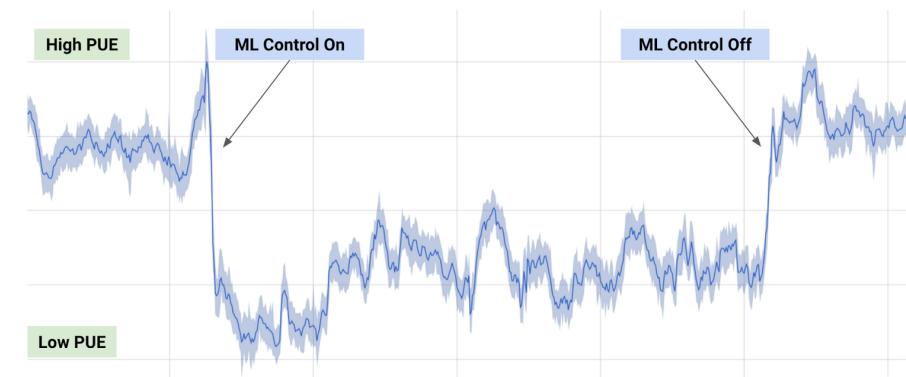
- Пример: крупный банк, больше миллиона клиентов
- Более 10 продуктов
- Каналы: sms, e-mail, звонок, уведомление в онлайн-банке, ...
- Увеличение LTV: up-sell/cross-sell
- Классический подход: сегментация клиентов
- Добавление персонализации: +13% дохода

Предсказание оттока

- Клиент ушёл, если не пользовался услугами больше N дней
- Модель может предсказать, кто уйдёт через 1-2 недели
- Для топ-5% запускается кампания по удержанию
- Классические подход: агрегированные данные и простые модели
- Анализ больших массивов данных и сложные нелинейные модели дают прирост от 10% до 20% по lift-10%

Автоматизация системы охлаждения

- Одна из главных компонент дата-центра — система охлаждения
- Результат работы системы сложным образом зависит от её параметров
- Необходимо быстро адаптироваться под изменение условий (нагрузка на серверы, погода)
- Все дата-центры разные — эвристические правила одного центра не работают в другом
- Машинное обучение позволило сократить затраты электричества на охлаждение на 40%



Выплавка стали

- Модель предсказывает, получится ли требуемый химический состав в результате плавки
- Сокращает расход ферросплавов на 5%
- Экономия до 23 млн. руб. в месяц
- Совместный проект Яндекса и Магнитогорского металлургического комбината

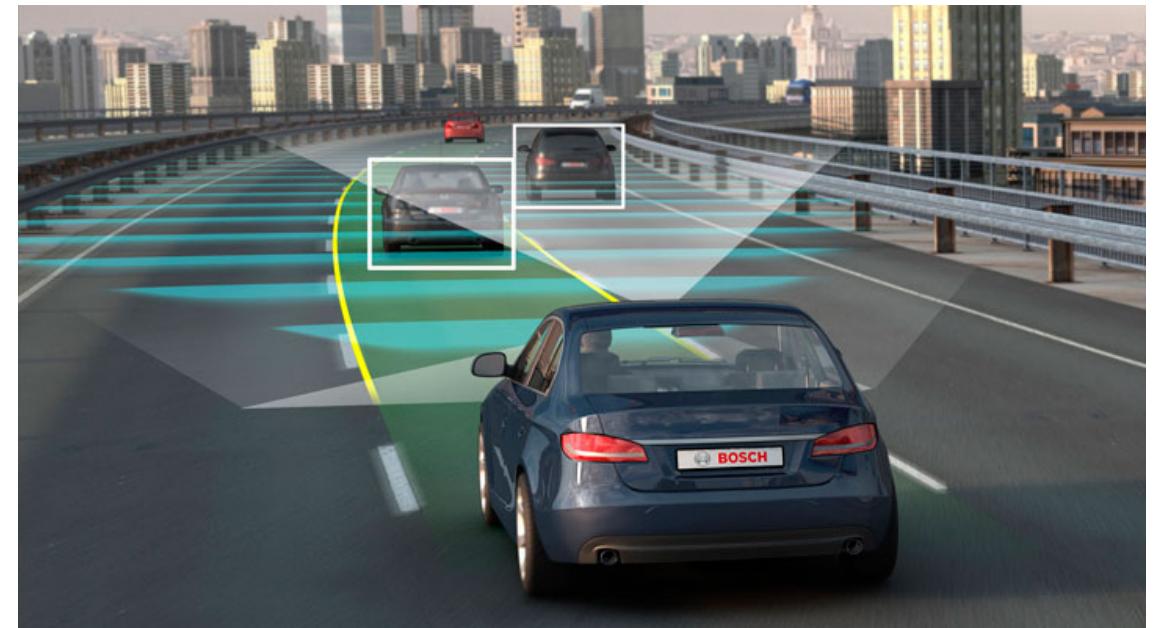


Машинное обучение в юриспруденции

- Поиск близких дел и материалов
- Проверка договоров, оптимизация правовых текстов
- Автоматизация принятия решений в суде
- Выявление проблем и трендов
- Правовые вопросы: обработка и передача больших данных, персональные данные, творчество, автопилоты

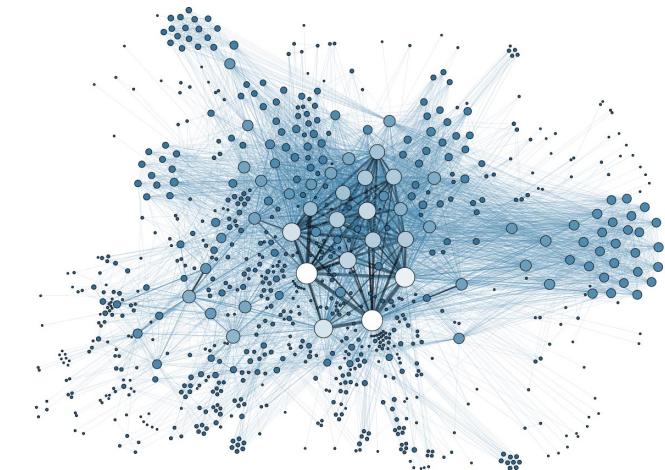
Экономические вызовы

- Создание новых отраслей экономики
 - Личные помощники
 - Персонализированная медицина
 - Гибкая сфера услуг
- Отмирание ряда профессий
 - Дальнобойщики
 - Водители такси
 - Пилоты самолётов
 - Операторы колл-центров



Машинное обучение в экономике и социологии

- Финансы и финтех: анализ транзакционных и поведенческих данных, обнаружение фрода
- Страхование: предсказание вероятности аварии по данным с датчиков автомобиля
- Оценка мнений и трендов: анализ данных социальных сетей, форумов, новостных ресурсов
- Data-driven принятие решений



Машинное обучение в HR

Машинное обучение в HR

- Поиск кандидатов и предсказание исхода собеседования
- Помощь при ротации
- Предсказание ухода сотрудника
- Анализ внутренних форумов, выделение жалоб

Предсказание увольнения

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences
49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences
37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other
33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences
27	No	Travel_Rarely	591	Research & Development	2	1	Medical
32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences
59	No	Travel_Rarely	1324	Research & Development	3	3	Medical

Предсказание увольнения

```
##          pred
## Attrition No Yes
##      No    167   15
##      Yes    11   18
```

```
[[1]]
[[1]]$accuracy
[1] 0.8767773

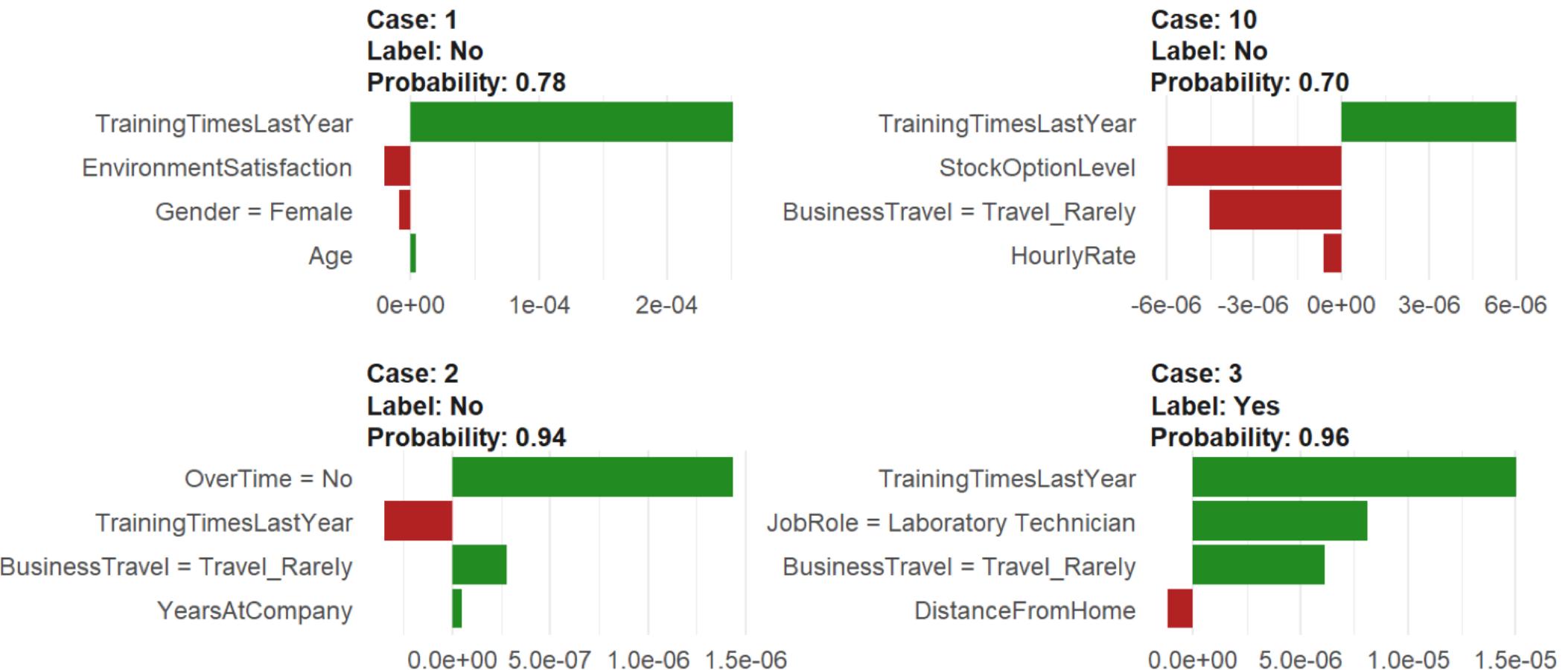
[[1]]$misclassification_rate
[1] 0.1232227

[[1]]$recall
[1] 0.6206897

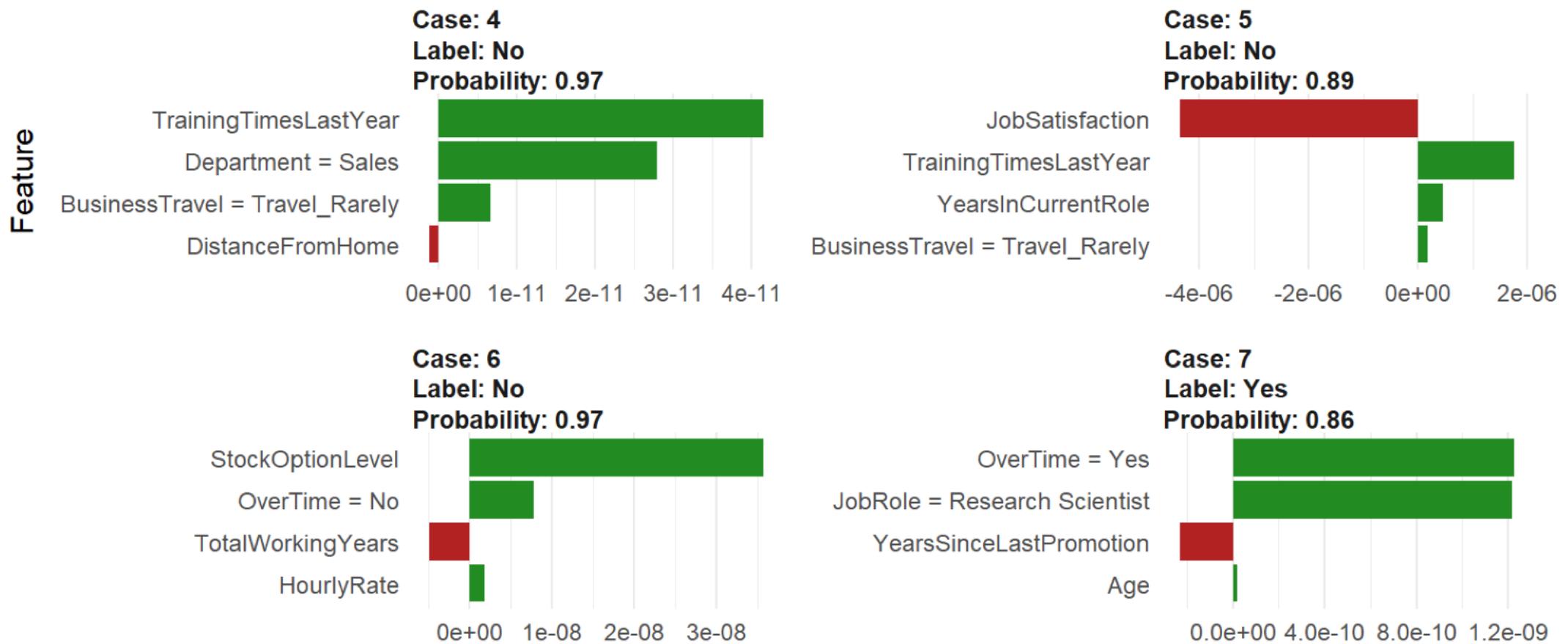
[[1]]$precision
[1] 0.5454545

[[1]]$null_error_rate
[1] 0.7914692
```

Предсказание увольнения



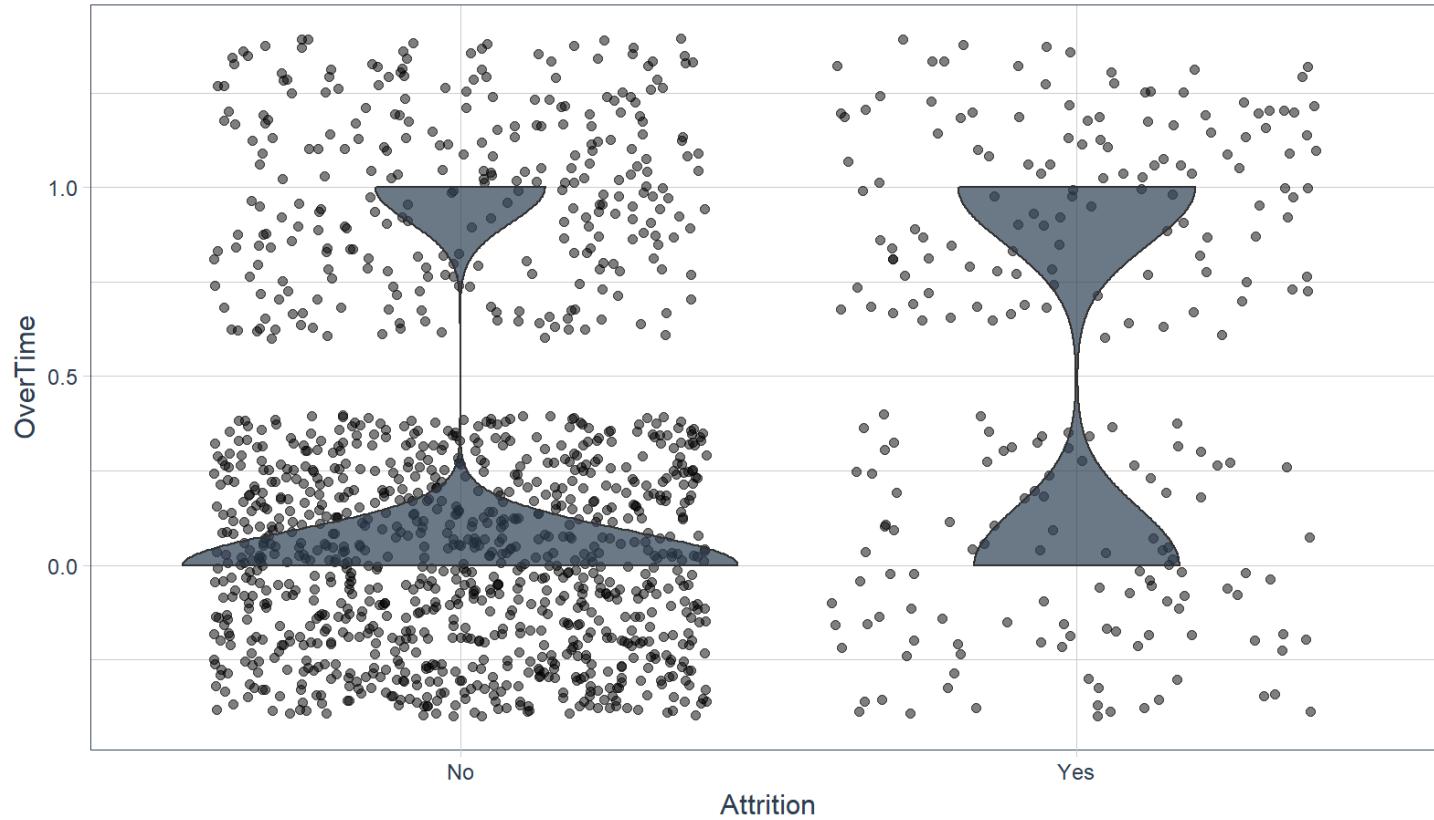
Предсказание увольнения



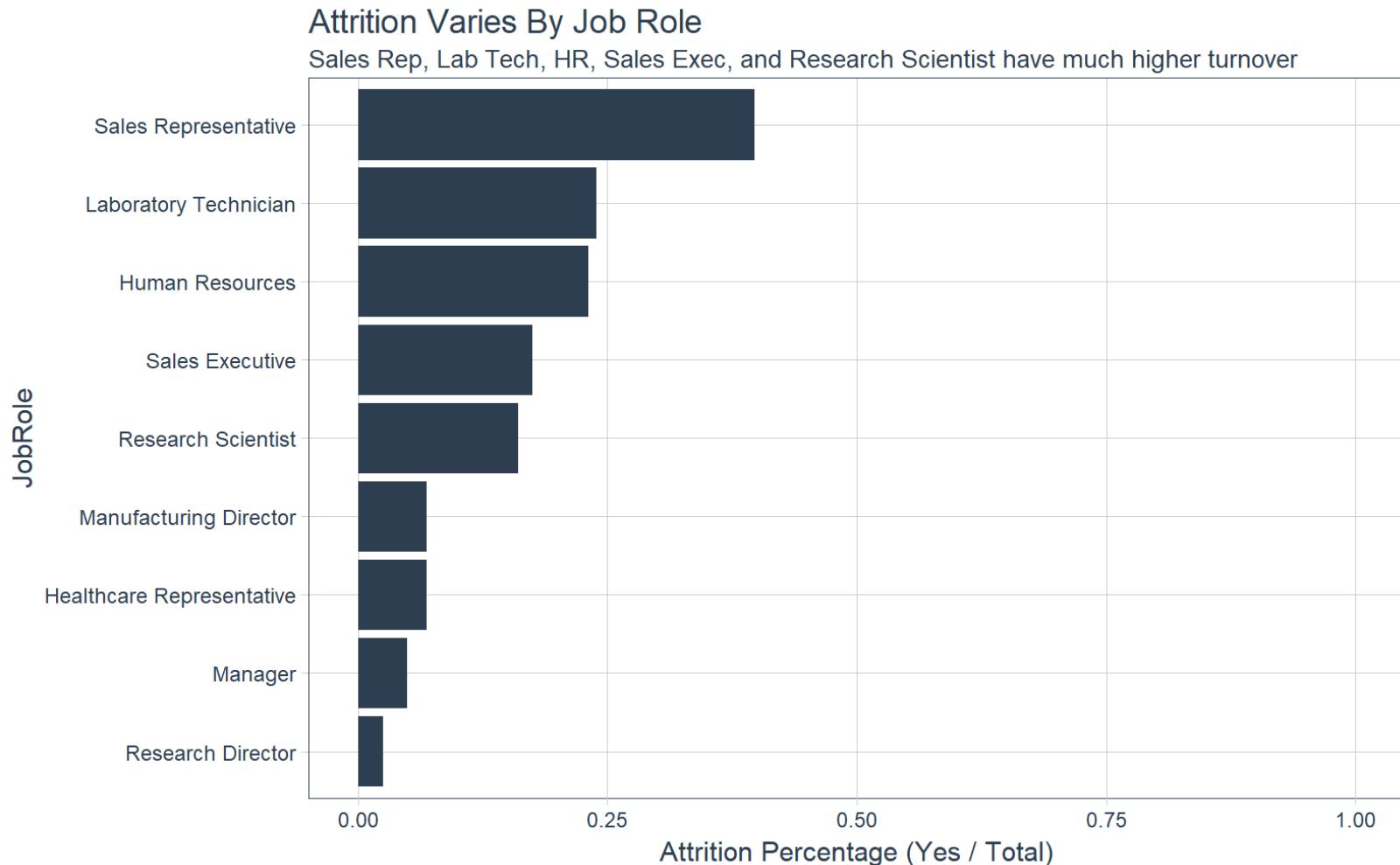
Предсказание увольнения

Prevalence of Over Time is Higher in Attrition = Yes

Suggests that increased overtime is related to higher attrition



Предсказание увольнения



Анализ отзывов от сотрудников

- Google: каждый год сотрудники проходят опрос, где сотрудники отвечают на вопросы о менеджерах
- Из отзывов можно выделить характеристики, присущие хорошим и плохим менеджерам

Анализ отзывов от сотрудников

What makes a good manager?

- Good at coaching employees
- Empowers teams without micro-managing
- Cares about their teams, supports work-life balance
- Productive and results orientated
- Good communicator
- Helps staff with career development
- Has a clear vision and strategy for their teams
- Has important technical skills, and can advise their teams on technical issues

Source: Google

Анализ отзывов от сотрудников

Signs that managers are struggling

- They have had a tough transition into the manager role, perhaps because they have been hired externally or without training support.
- They lack a consistent approach to performance management. They may be unaware of Google's performance management approach.
- They don't support career development, or may equate career development with promotion.
- They expect authority and respect as a result of their position, rather than by earning it.
- They spend too little time on managing and communicating

Source: Google

Резюме

- Современные методы машинного обучения позволяют генерировать картины и тексты, воспроизводя уже существующие стили
- Рекомендательные системы — новый способ персонализации предложений и контента
- Автоматизация процессов за счёт обучения на больших объёмах данных
- HR-аналитика — поиск нетривиальных закономерностей