

Программа «LegalTech директор»

# Технологии машинного обучения и обработки естественного языка для LegalTech

*Воронцов Константин Вячеславович*

- МФТИ, руководитель лаборатории Машинного интеллекта ●
  - ФИЦ ИУ РАН ● Сбербанк ● ШАД Яндекс ●

[k.v.vorontsov@phystech.edu](mailto:k.v.vorontsov@phystech.edu)

# Машинное обучение в LegalTech

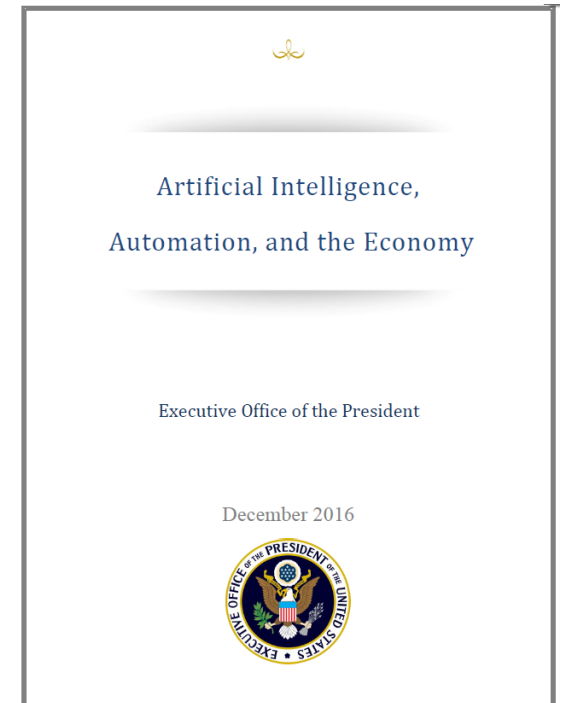
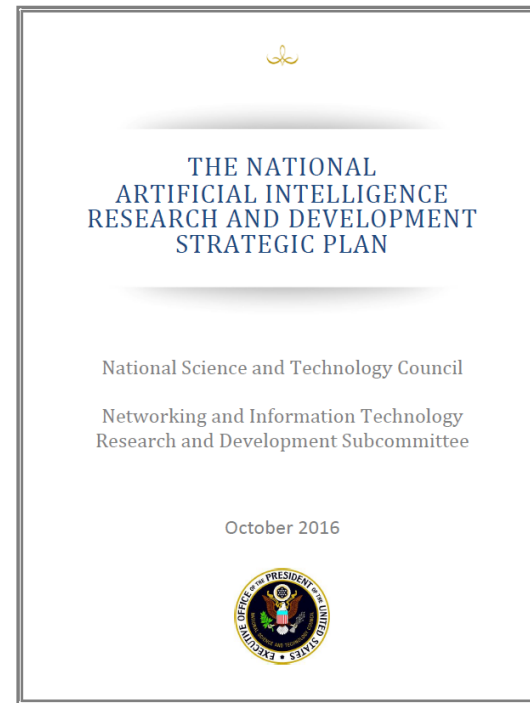
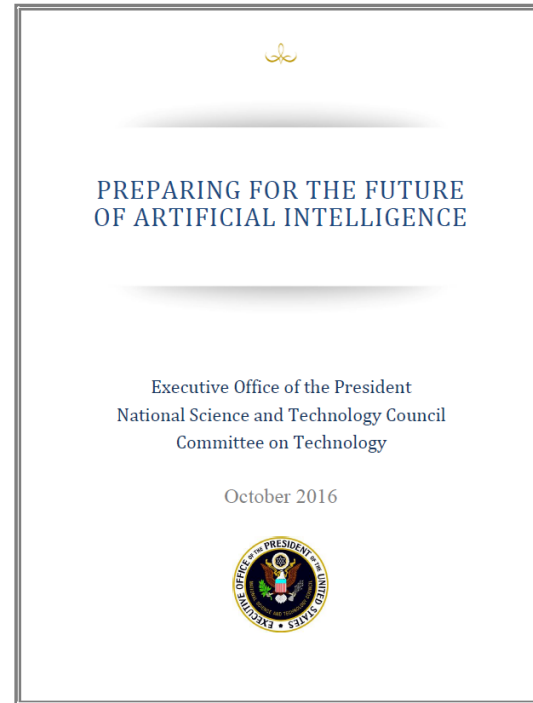
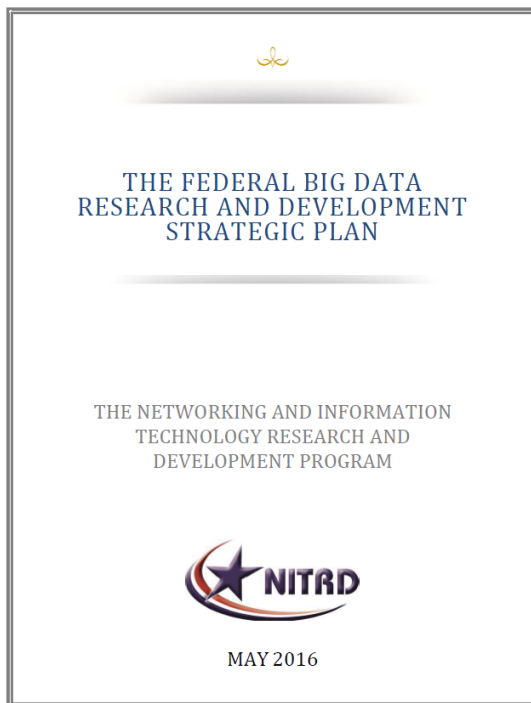
1. Задачи, возможности и ограничения машинного обучения
  - Бум искусственного интеллекта и нейронных сетей
  - Постановки задач и терминология и методология машинного обучения
  - Методология машинного обучения
2. Задачи и технологии обработки естественного языка
  - Классификация задач и технологий анализа текстов
  - Разведочный информационный поиск
  - Примеры задач анализа текстов
3. LegalTech: задачи, технологии, сервисы

«Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, *искусственном интеллекте* и *машинном обучении*» (2016)

Клаус Мартин Шваб,  
президент Всемирного  
экономического форума



# Отчёты Белого дома США, май-октябрь 2016



«Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future»

# Бум искусственного интеллекта

**1997:** IBM Deep Blue обыграл чемпиона мира по шахматам

**2005:** Беспилотный автомобиль: DARPA Grand Challenge

**2006:** Google Translate – статистический машинный перевод

**2011:** 40 лет DARPA CALO привели к созданию Apple Siri

**2011:** IBM Watson победил в ТВ-игре «Jeopardy!»

**2011–2018:** ImageNet: 25% → 2,5% ошибок против 5% у людей

**2015:** Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана

**2016:** DeepMind, OpenAI: динамическое обучение играм Atari

**2016:** Google DeepMind обыграл чемпиона мира по игре го

**2017:** OpenAI обыграл чемпиона мира по компьютерной игре Dota 2



# Глубокие нейронные сети обеспечили прорыв в компьютерном зрении

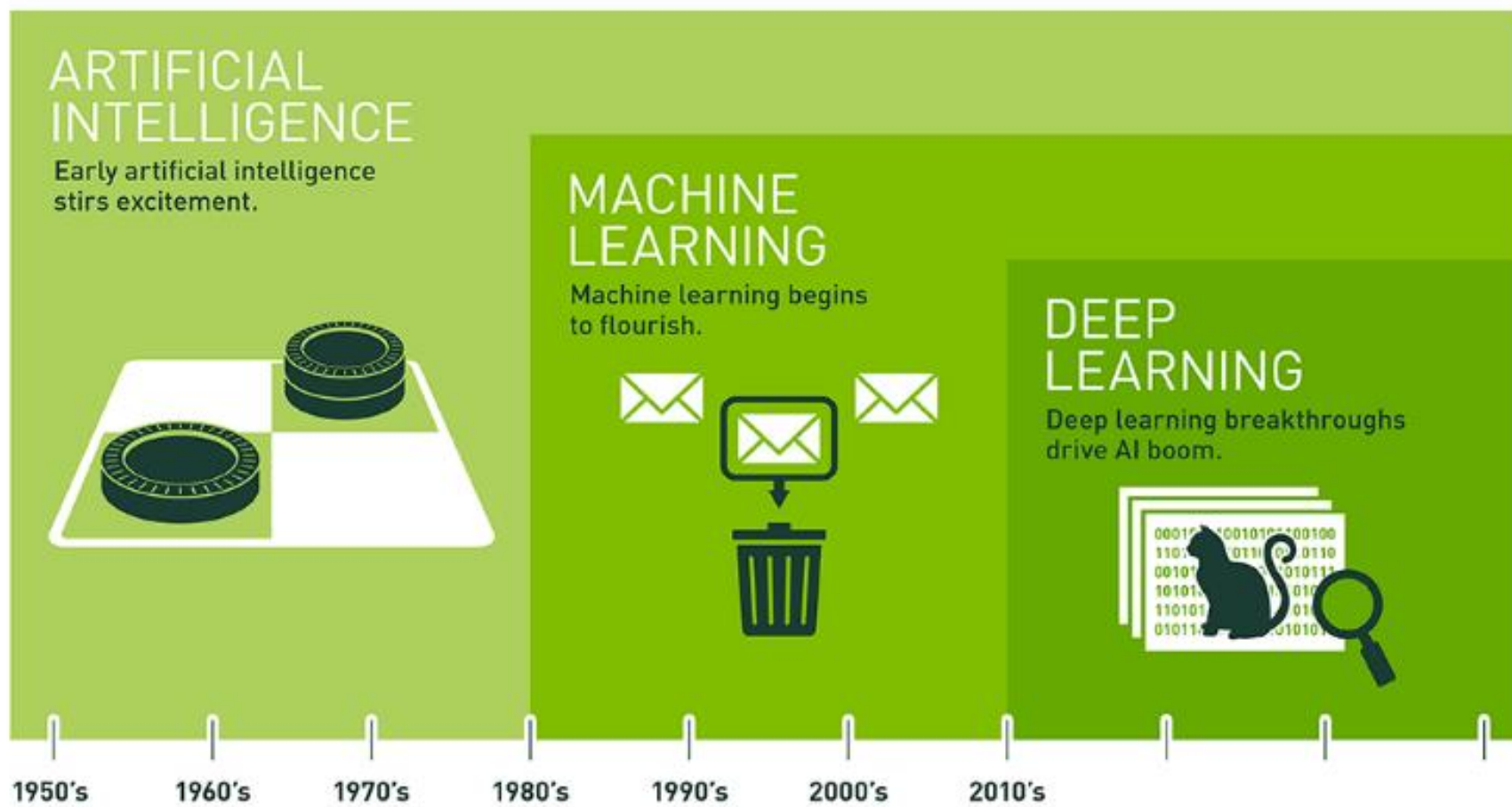
**ImageNet:** открытая выборка 15М размеченных изображений



**Google:** Распознавание кадров с котами на видео из Youtube



# Эволюция искусственного интеллекта

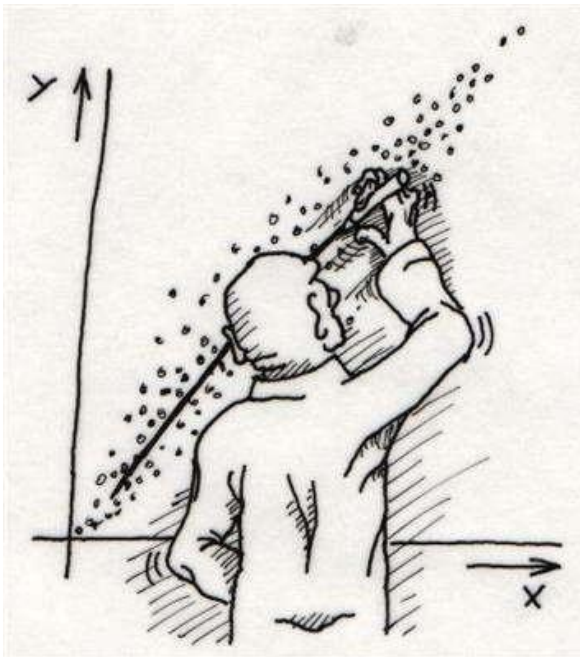


*Глубокое обучение  
– одна из новейших  
технологий  
машинного  
обучения*

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Машинное обучение (Machine Learning, ML)

- одна из ключевых информационных технологий будущего
- наиболее успешное направление *искусственного интеллекта*, вытеснившее экспертные системы и инженерию знаний
- проведение функции через заданные точки в сложно устроенных пространствах
- математическое моделирование в условиях, когда знаний мало, данных много
- тысячи различных методов и алгоритмов
- около 100 000 научных публикаций в год





# Основная задача машинного обучения

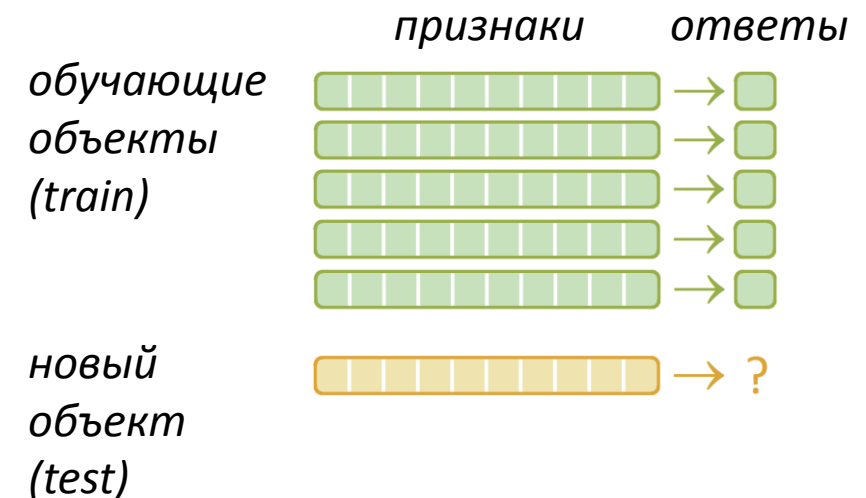
## Этап №1 – обучение с учителем

- **На входе:**  
*данные* – выборка прецедентов «*объект* → *ответ*»,  
каждый объект описывается набором *признаков*
- **На выходе:**  
модель, предсказывающая ответ по объекту

Если нет данных,  
то нет  
и машинного  
обучения

## Этап №2 – применение

- **На входе:**  
*данные* – новый объект
- **На выходе:**  
предсказание ответа на новом объекте



# Примеры задач машинного обучения

- **Кредитный скоринг:**

объект – данные о заёмщике

ответ – вероятность дефолта, решение по кредиту



- **Информационный поиск в Интернете:**

объект – данные о паре «запрос и документ»

ответ – оценка релевантности документа запросу



- **Рекомендательные системы в Интернете / TV:**

объект – данные о паре «пользователь, товар / фильм»

ответ – оценка вероятности покупки / просмотра

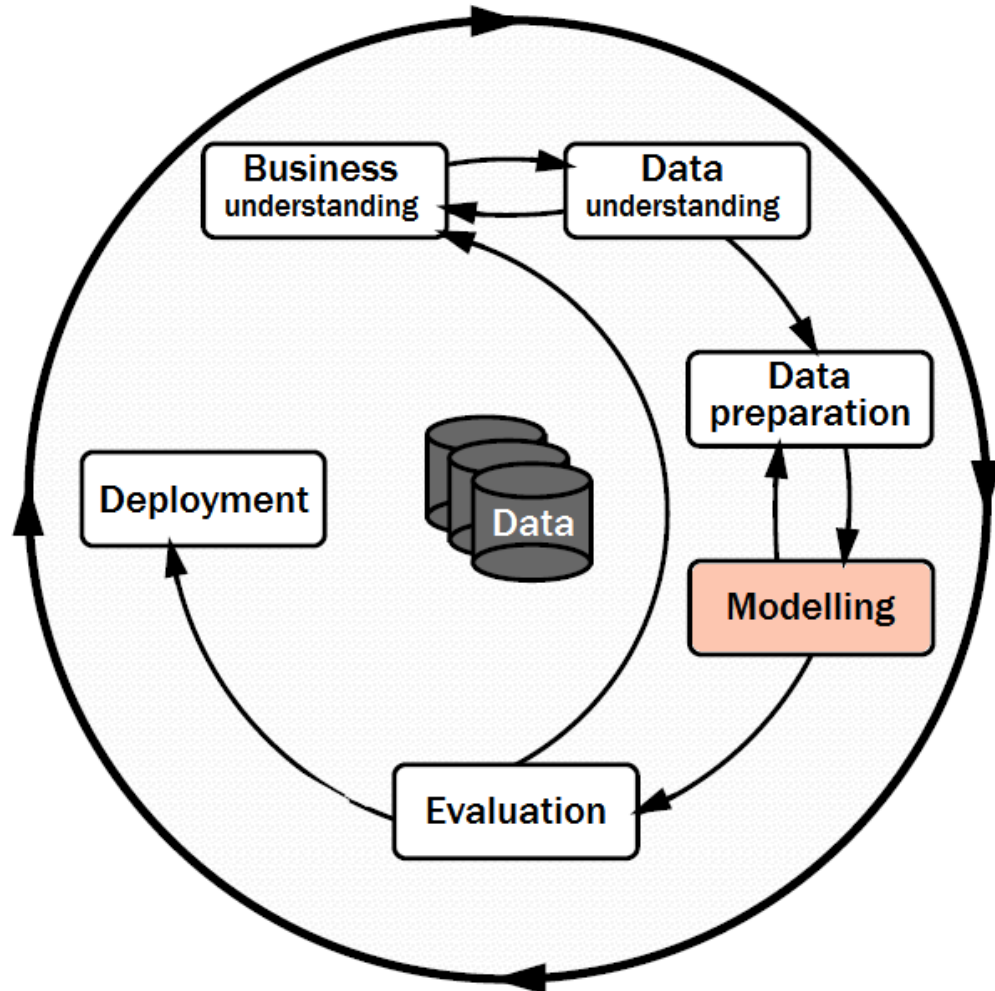


# Примеры задач ML в LegalTech

- **Предсказание результатов судебных процессов:**  
объект – описание дела: документы, записи о событиях  
ответ – вероятность выиграть дело
- **Информационный поиск:**  
объект – описание дела или вопрос на естественном языке  
ответ – ранжированный список релевантных НПА и/или схожих дел
- **Рекомендательный сервис:**  
объект – пара «описание дела, профиль юриста / юрфирмы»  
ответ – список консультантов, ранжированный по релевантности
- **Автоматическая генерация ответов на вопросы:**  
объект – текст вопроса на естественном языке  
ответ – текст ответа, включая фрагменты НПА

# Этапы решения задач ML/DS/AI

CRISP-DM: Cross Industry Standard Process for Data Mining (1999)



- понимание бизнес-задач
- понимание данных
- предобработка данных
- инженерия признаков
- построение моделей
- оптимизация параметров
- контроль переобучения
- (кросс-)валидация решения
- внедрение и эксплуатация

# Открытые данные для ИИ

- **Выгоды открытых данных**
  - быстрое выявление центров компетенций, подбор кадров
  - быстрое выявление наилучших решений и «потолка качества»
  - формирование проф.сообществ, обучение на реальных кейсах
  - популяризация научных знаний в любых областях, где есть данные
- **Конкурсы анализа данных**
  - [www.NetflixPrize.com](http://www.NetflixPrize.com) (2006-2009) – первый крупный конкурс, \$1 млн.
  - [www.kaggle.com](http://www.kaggle.com) – самая известная платформа
  - [www.FakeNewsChallenge.org](http://www.FakeNewsChallenge.org) – один из последних конкурсов
- **[DataRing.ru](http://DataRing.ru)** – отечественная конкурсная платформа
  - консалтинг по подготовке данных и условий конкурса
  - очистка, отбор, агрегирование, деперсонификация данных

# Что должен уметь *инженер* в области Data Science

- Работать с сырыми данными в любых форматах
- Визуализировать, понимать, очищать, преобразовывать данные
- Выбирать инструменты и методы под задачу
- Строить признаки/архитектуры (feature/architecture engineering)
- Оценивать решения в соответствии с бизнес-целями
- Доводить пилотное решение до внедрения (production)
- Ходить по кругу CRISP-DM

# Что должен уметь *менеджер* в области Data Science

- Видеть возможности применения машинного обучения
- Ставить задачи в виде «Д-Н-К» (Дано-Найти-Критерий)
- Разбираться в методах на уровне «возможности–ограничения»
- Организовывать бизнес-процессы для сбора чистых данных
- Организовывать открытые конкурсы анализа данных
- Запускать пилотные проекты
- Знать экспертное сообщество
- Адекватно оценивать сложность задач и трудозатраты

# Машинное обучение в LegalTech

1. Задачи, возможности и ограничения машинного обучения
  - Бум искусственного интеллекта и нейронных сетей
  - Постановки задач и терминология и методология машинного обучения
  - Методология машинного обучения
2. Задачи и технологии обработки естественного языка
  - Классификация задач и технологий анализа текстов
  - Примеры задач анализа текстов
  - Разведочный информационный поиск
3. LegalTech: задачи, технологии, сервисы



# Задачи обработки естественного языка

## Вспомогательные лингвистические задачи:

Morphological Parsing  
Part-of-speech Tagging  
Syntax Parsing  
Semantic Role Labeling  
Named Entity Recognition  
Automatic Term Extraction  
Word Sense Disambiguation  
Paraphrase Identification  
Topic Modeling  
Word Embedding  
Intent Recognition

## Конечные бизнес-задачи:

Conversational Intelligence  
Machine Translation  
Question Answering  
Information Retrieval  
Sentiment Analysis  
Text Summarization  
Text Segmentation  
Text Classification  
Text Clustering  
Fact Extraction  
Ontology Learning



# Стек технологий NLP (см. также [nlpub.ru](http://nlpub.ru))

- Морфологический анализ и лемматизация (morphological parsing, lemmatization)
- Синтаксический анализ (syntax parsing)
- Автоматическое выделение терминов (automatic term extraction)
- Распознавание именованных сущностей (named entity recognition)
- Сегментация текста (text segmentation)
- Классификация текстов (text classification)
- Кластеризация текстов (text clustering)
- Тематическое моделирование (topic modeling)
- Семантические векторные представления слов (word embedding)
- Семантический анализ и построение онтологий (ontology learning)
- Аннотирование и суммаризация (text summarization)
- Обучаемое ранжирование (learning to rank)
- Информационный поиск (information retrieval)
- Ответы на вопросы, машинный перевод, чат-боты (sequence-to-sequence)

# Выделение смысла? Понимание речи?

- Поиск «смысла» бессмысленен
- Что такое «понимание», не понятно
- Бизнес, технологии и математика работают только с чётко определяемыми понятиями и чётко поставленными задачами
- Задача чётко поставлена, если для неё описано **ДНК**:  
«что **Д**ано – что **Н**айти – **К**ритерий качества решения»
- Измеримый критерий появляется, когда цели прагматичны:
  - автоматизация рутинных операций
  - повышение производительности труда
  - снижение издержек

# Условная классификация задач NLP

- По структуре входов-выходов «чёрного ящика»:
  - классификация, предсказательное моделирование:  
**ВХОД** – текст, **ВЫХОД** – число
  - информационный поиск:  
**ВХОД** – текст, **ВЫХОД** – ранжированный список документов
  - преобразование и синтез текста:  
**ВХОД** – текст, **ВЫХОД** – текст
- По критерию качества и положению в цепочке обработки данных:
  - бизнес-задачи
  - вспомогательные задачи компьютерной лингвистики
- По уровням анализа текста (пирамида NLP)

# Классификация, предсказательные модели

## **Дано:**

- текстовый документ (e-mail, web-страница)

## **Найти:**

- один из двух классов: спам / не-спам

## **Критерий:**

- AUC, чувствительность и специфичность

Модель классификации строится по обучающей выборке,

Основная подзадача: преобразовать текст в векторное признаковое описание фиксированной размерности.

# Задача классификации отзывов/обращений

## **Дано:**

- текст отзыва или обращения клиента

## **Найти:**

- класс: куда маршрутизировать запрос / о какой проблеме сообщает клиент или сотрудник

## **Критерий:**

- AUC, чувствительность и специфичность (для многоклассовой классификации)

Модель классификации строится по обучающей выборке,

Основная подзадача: преобразовать текст в векторное признаковое описание фиксированной размерности.

# Задача анализа тональности (Sentiment Analysis)

## **Дано:**

- текст отзыва или обращения клиента

## **Найти:**

- оценку тональности отзыва в целом, от -1 до +1

## **Критерий:**

- точность определения тональности на размеченных данных

Модель классификации строится по обучающей выборке,

Используются словари тональных слов.

# Задача ранжирования поисковой выдачи (Learning-to-Rank, Information Retrieval)

## **Дано:**

- пара «короткий запрос, документ»

## **Найти:**

- оценка релевантности документа запросу

## **Критерий:**

- точность и полнота поиска по ассессорским данным
- качество ранжирования

Надо придумывать признаки, распознавать и исправлять опечатки, учитывать словоформы, синонимы, парафразы.



# Оценки качества информационного поиска

Precision — доля релевантных среди найденных

Recall — доля найденных среди релевантных

$$P = \frac{TP}{TP + FP} \text{ — точность (precision)}$$

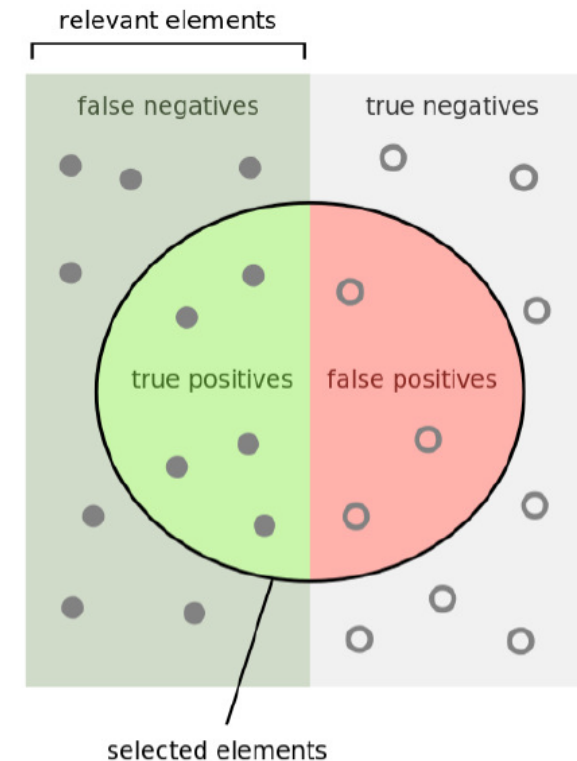
$$R = \frac{TP}{TP + FN} \text{ — полнота, (recall)}$$

$$F_1 = \frac{P + R}{2PR} \text{ — F1-мера}$$

TP (true positive) — найденные релевантные

FP (false positive) — найденные нерелевантные

FN (false negative) — ненайденные релевантные



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

# Машинный перевод (Machine Translation)

## **Дано:**

- текст на одном языке

## **Найти:**

- его перевод на другой язык

## **Критерий:**

- близость к профессиональному переводу, число исправлений

Обучающие данные: большой корпус параллельных текстов, частично с выравниванием, двуязычные словари.

# Суммаризация и аннотирование (Summarization)

## **Дано:**

- документ или подборка документов

## **Найти:**

- краткое содержание (реферат)

## **Критерий:**

- точность соответствия (как правило, нескольким) рефератам, написанным людьми (метрики ROUGE, BLUE)

Особенности задачи: надо учитывать словоформы, синонимы, парафразы; надо выбирать самое важное, но без повторов

# Ответы на вопросы (Question Answering)

## **Дано:**

- текст вопроса

## **Найти:**

- текст ответа на поставленный вопрос

## **Критерий:**

- точность выделения фразы ответа на размеченной выборке пар «вопрос - текст-с-ответом»

Обучающие данные: коллекция пар «вопрос – ответ»

# Разговорный интеллект (Conversational Intelligence, chatbots)

## **Дано:**

- текст диалога бота с человеком

## **Найти:**

- следующую реплику бота

## **Критерий:**

- тест Тьюринга: человек-судья не может отличить собеседника-человека от собеседника-бота
- в приложениях: доля случаев, когда потребность клиента была удовлетворена / когда оператор принял подсказку бота

Обучающие данные: коллекция диалогов операторов с клиентами

# Примеры задач анализа текстов

#1: Выделение значений параметров

#2: Тегирование звонков в контакт-центр

#3: Классификация отзывов клиентов по категориям

#4: Кластеризация отзывов по темам

#5: Тематическая сегментация записей разговоров контактного центра

#6: Разведочный информационный поиск

# #1: Выделение значений параметров

**Цель:** автоматизировать анализ конкурсной документации по госзакупкам:

**Задача:** находить и выделять в текстах значения параметров:

- Дата начала выполнения работ
- Дата окончания действия контракта
- Размер обеспечения заявки
- Наличие аванса

**Критерий:** точность распознавания размеченных полей

**Метод:** фиксированные или обучаемые правила

**Результат:** точность близка к 100%

# #1: Сложность задачи – разнообразие фраз

**Пример.** Встречающиеся способы описания *даты начала работ*:

...контракт вступает в силу с момента заключения контракта и действует по **31 марта 2019 г.** включительно

...срок выполнения работ:

начало: 15 декабря 2017 года; окончание – **31 декабря 2018 года**

...настоящий договор действует до исполнения обязательств сторонами, но не позднее **31 декабря 2019 года**

...срок действия которых истекает не ранее **25 сентября 2019 г**

...срок оказания услуги: с момента заключения договора (но не ранее 09.01.2017 г.) по **31.12.2017 г.**



## #2: Тегирование звонков в контакт-центр

**Цель:** оценивание результативности маркетинговых акций,  
оценивание рекламных площадок,  
оценивание качества работы операторов

**Задача:** определение намерений клиента и результата разговора

**Критерий:** точность, полнота, F1-мера по размеченной выборке записей разговоров

**Методы:** логистическая регрессия с отбором признаков,  
кросс-валидация со стратификацией классов

**Результат:** F1-мера от 50% до 90% в зависимости от класса

## #2: Звонки в автосалоны

### Классификация разговоров:

- запись на тех. обслуживание
- автомобиль в кредит
- ВИЗИТ В АВТОСАЛОН
- trade-in

### Вспомогательная задача:

- определение марки и модели автомобиля

Категория	Precision	Recall	F-1 score	Accuracy
Договор о встрече	53,6%	89%	66,9%	75,5%
Trade-In	46,5%	58,6%	51,9%	80,7%
Тех. Обслуживание	66,6%	92,6%	77,5%	93,2%
Марка автомобиля	86,6%	93,7%	90%	81,7%
Модель автомобиля	66,3%	78%	71,7%	55,8%

# #2: Звонки в риэлторские компании

## Классификация разговоров:

- договорённость о встрече
- договорённость о перезвоне
- готовность к оплате
- ипотека

## Вспомогательная задача:

- параметры объекта недвижимости
- наличие нецензурной лексики

Категория	Precision	Recall	F-1 score
Договор о встрече	45%	60%	51%
Договор о перезвоне	79%	73%	76%
Ипотека	61%	66%	64%
Квартира в аренду	66%	80%	73%
Покупка квартиры	87%	90%	88%

# #2: Интерпретируемый отбор признаков

## Пример 1: Авто в кредит

- *взнос кредит кредитный ставка процент кредитование процентный условие программа рассчитать посчитать платёж встреча банка сожаление клиент зачёт сдавать визитка подать сервис акция хотеться ожидание срок смочь знать покупка самый выбрать отправить записаться брать чёрный встретиться новое обсудить*

## Пример 2: Ипотека

- *ипотека банк компания справка площадь отделка условие ндфл документ втб отдел собственность принцип проблема одобрение апартамент история номер комнатка контакт сбербанк этаж объект станция сдача адрес дело улица размер знать ремонт лицо ставка планировка новое консультация процесс смска недвижимость координата*

# #3: Классификация отзывов по категориям

**Цель:** анализ отзывов потребителей по каналам обратной связи (горячая линия, VK, telegram, mail, форум...)

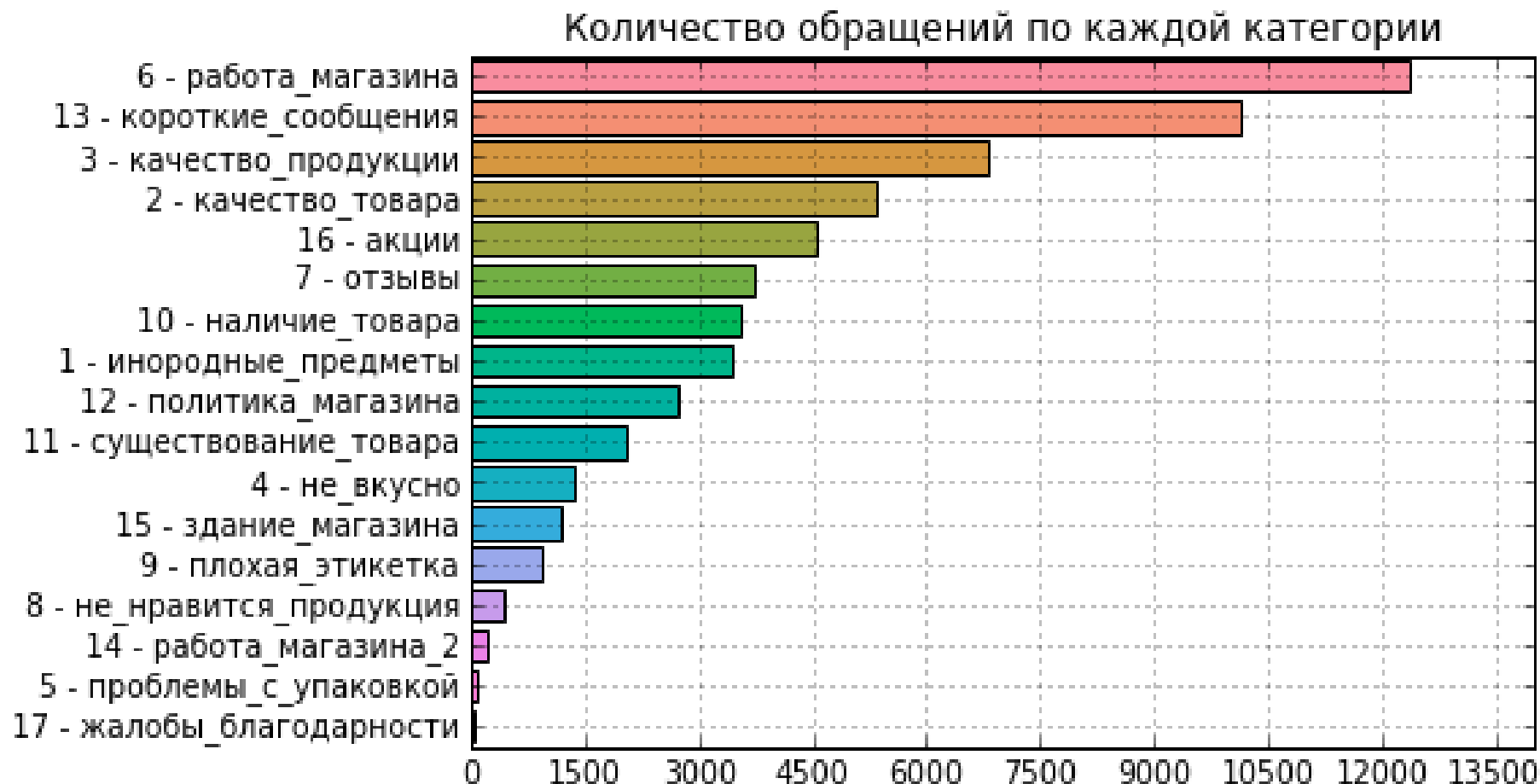
**Задача:** классификация отзывов по *17 известным категориям*

**Критерий:** точность, полнота, F1-мера по размеченной выборке

**Методы:** логистическая регрессия,  
градиентный бустинг,  
отбор признаков,  
кросс-валидация со стратификацией классов

**Результат:** F1-мера от 50% до 90% в зависимости от класса

# #3: Сложность задачи – несбалансированность классов

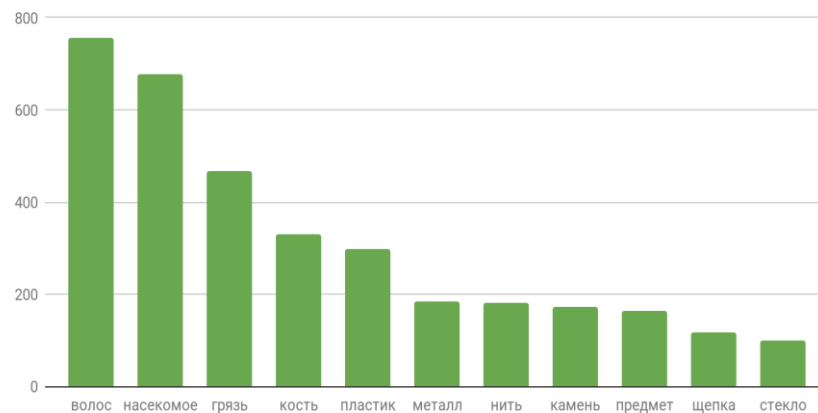


# #3: Примеры отчётов

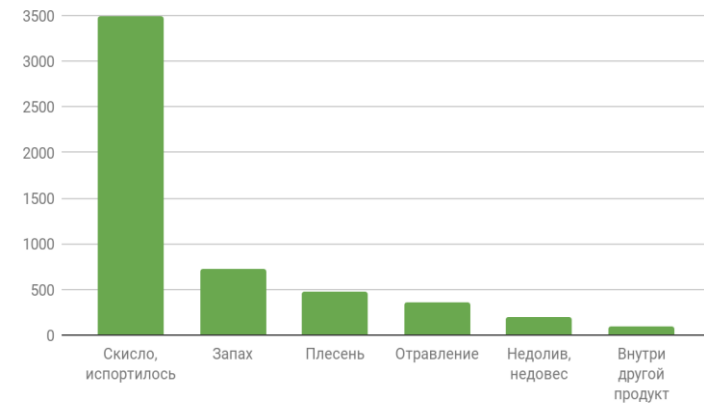
Отчёты строятся

- по категориям проблем
- по времени
- по каналам коммуникации

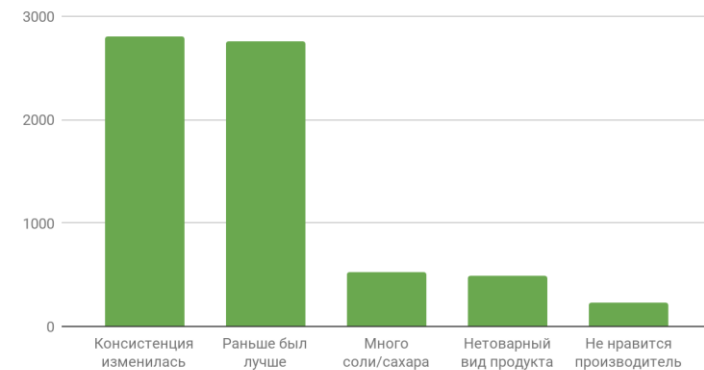
Инородные предметы



Качество товара



Качество продукции



# #4: Кластеризация отзывов по темам

- Цель:** анализ отзывов пользователей приложения для оформления заказов в ресторанах быстрого питания
- Задача:** кластеризация (тематическое моделирование) отзывов по кластерам (темам), которые *заранее не известны*
- Критерий:** точность, полнота, F1-мера по размеченной выборке
- Методы:** тематическое моделирование (BigARTM), анализ тональности
- Результат:** F1-мера улучшилась от 61% (простейшая модель) до 81% (+модальности +частичное обучение)



# #4: Результат тематического моделирования

<b>Deals</b> <ul style="list-style-type: none"> <li><i>deal, coupon, menu, offer, drink, meal, sandwich, price, option, beverage</i></li> </ul>	<b>Punchcard</b> <ul style="list-style-type: none"> <li><i>punchcard, punch, coupon_code, show, save, free, coffee, price, drink, reward</i></li> </ul>	<b>Competitor mentions</b> <ul style="list-style-type: none"> <li><i>restaurant, glitch, pass, log_in, get_passed, download, waste, crash, uninstall</i></li> </ul>
<b>Ordering</b> <ul style="list-style-type: none"> <li><i>menu, order, mobile, card, drink, meal, price, discount, purchase, cashier</i></li> </ul>	<b>Stability &amp; Performance</b> <ul style="list-style-type: none"> <li><i>buggy, load, freeze, slow, terrible, fix, install, down, version, app</i></li> </ul>	<b>Location</b> <ul style="list-style-type: none"> <li><i>location, gps, place, gps_location, address, map, place, store, work, find_location</i></li> </ul>
<b>UX Issues</b> <ul style="list-style-type: none"> <li><i>crash, usability, trouble, terrible, password, install, review, login, get, uninstall</i></li> </ul>	<b>Store Issues (Crew)</b> <ul style="list-style-type: none"> <li><i>employee, work, train, fix_issues, scan, uninformed, problem, use_app, customer, app</i></li> </ul>	<b>Store Issues (Restaurant)</b> <ul style="list-style-type: none"> <li><i>order, food, lunch, update, favourite, open, place, menu, card, mobile</i></li> </ul>
<b>Registration</b> <ul style="list-style-type: none"> <li><i>register, log_in, get_passed, download, install, account, load, password, notification</i></li> </ul>	<b>Login</b> <ul style="list-style-type: none"> <li><i>log, glitch, pass, log_in, get_passed, waste, crash, uninstall, account, sign_in</i></li> </ul>	<b>Feature Request</b> <ul style="list-style-type: none"> <li><i>find, price, device, calculator, deal, location, permission, password, new, app</i></li> </ul>
<b>New version/Release Issue</b> <ul style="list-style-type: none"> <li><i>update, version, new, change, bad, work, release, new_app, bug, memory</i></li> </ul>	<b>Other</b> <ul style="list-style-type: none"> <li><i>food, lovin_it, love, like, wonderful, awesome, delicious, work, save_money</i></li> </ul>	<b>These notable keywords were automatically detected by the topic model for each top-level category. No manual filtering was done.</b>

# #5: Тематическая сегментация записей разговоров контактного центра

- Цель:** мониторинг качества работы операторов, выявление лучших практик, генерация подсказок операторам
- Задача:** разбиение разговора на короткие тематические сегменты, построение графа переходов между темами
- Критерий:** качество тематической сегментации
- Методы:** выделение терминов, синтаксический анализ, тематическое моделирование (BigARTM)
- Результат:** качество сегментации (доля правильно выделенных сегментов) возросло от 40% до 75%

# #5: Тематическая сегментация

(цвета – темы, подчёркивание – ассессорская разметка)

Оформление заявки	Индивидуальный подход	Решение банка	Доставка
Бонусная программа	Бесплатная доставка/оформление		

вот на данный момент я звоню предлагаю только составить заявку чтобы банк изучил вашу кредитную историю и подобрал под вас индивидуальный тарифный план после чего на ваш мобильный поступит уведомление в котором будет указано каким образом в случае положительного ответа будут доставлены бумаги у нас есть два способа доставки это либо курьерская доставка либо заказным письмом почтой России

ну вот бонусы значит на все абсолютно покупки один процент а если вы совершаете покупки у банка будет полный перечень магазинов у вас в личном кабинете до тридцати процентов бонусов можете то есть вот две тысячи что то купили а ориентировочно шестьсот вернулось вам на это уже плюсов согласитесь что это довольно таки это одна покупка вот так и хочу сказать что вы абсолютно ничего не теряетесь соглашаясь оформить заявку ничего за что не платите потому что вам карту выпускают доставляют абсолютно бесплатно вам либо представитель банка привозит либо по почте она приходит

# #6: Разведочный информационный поиск

- Цель:** поиск документов по смыслу, а не по ключевым словам, разведочный информационный поиск
- Задача:** многокритериальное тематическое моделирование, ранжирование документов по тематическому сходству
- Критерий:** точность и полнота поиска по размеченной выборке (коллекции habrahabr.ru и techcrunch.com)
- Методы:** выделение терминов (TopMine), тематическое моделирование (BigARTM)
- Результат:** точность и полнота поиска возросли с 68% до 93%, автоматизация задач поиска, занимающих 20-60 минут

# #6: Разведочный информационный поиск

## Пример разведочного запроса

IBM Watson — суперкомпьютер фирмы IBM, оснащённый вопросно-ответной системой искусственного интеллекта, созданный группой исследователей под руководством Дэвида Феруччи. Его создание — часть проекта DeepQA. Основная задача Уотсона — понимать вопросы, сформулированные на естественном языке, и находить на них ответы в базе данных. Назван в честь основателя IBM Томаса Уотсона.

IBM Watson представляет собой когнитивную систему, которая способна понимать, делать выводы и обучаться. Она также позволяет преобразовывать целые отрасли, различные направления науки и техники. Например, предсказывать появление эпидемий или возникновения очагов природных катастроф в различных регионах, вести мониторинг состояния атмосферы больших городов, оптимизировать бизнес-процессы, узнавать, какие товары будут в тренде в ближайшее время.

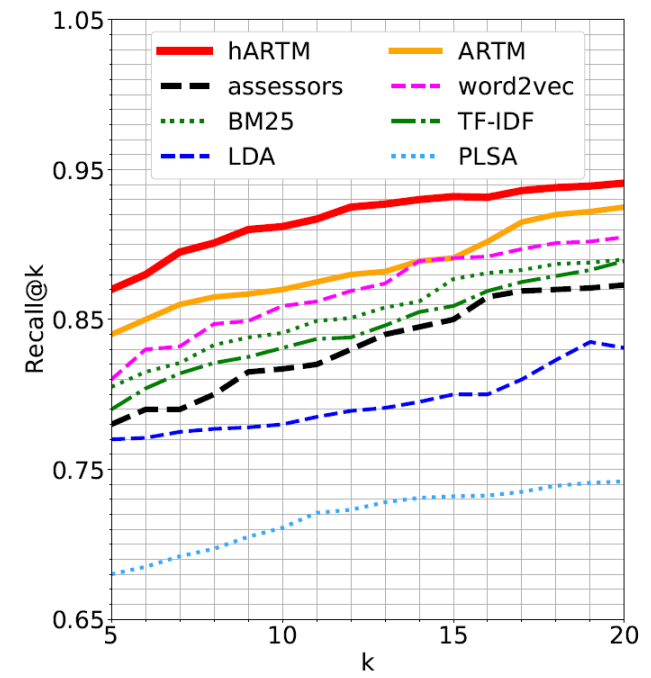
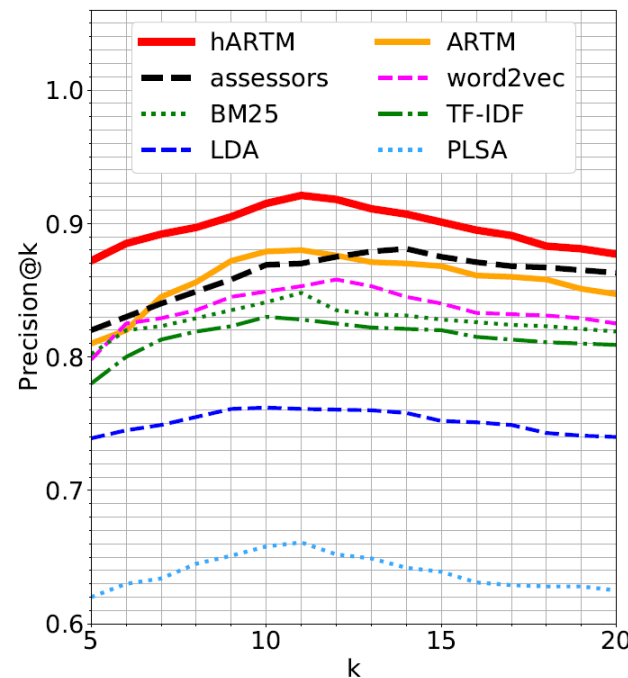
... ..

# #6: Разведочный информационный поиск

- Длинные запросы (1 стр. А4)
- 100 запросов
- 3 ассессора на каждый запрос
- 30 минут в среднем на запрос
- Разметка на Яндекс.Толока
- Коллекции техно-новостей



**Результат:** *точность* (precision) и *полнота* (recall) поиска



# Машинное обучение в LegalTech

1. Задачи, возможности и ограничения машинного обучения
  - Бум искусственного интеллекта и нейронных сетей
  - Постановки задач и терминология и методология машинного обучения
  - Методология машинного обучения
2. Задачи и технологии обработки естественного языка
  - Классификация задач и технологий анализа текстов
  - Разведочный информационный поиск
  - Примеры задач анализа текстов
3. LegalTech: задачи, технологии, сервисы

# Что в LegalTech не является AI/ML



Конструктор документов



Конструктор документов



Смарт-контракты



Автоматизация рутины в юридических фирмах



Быстрая регистрация торговых марок



Учёт судебных дел, доступ к данным госорганов



Проверка документов на соответствие закону



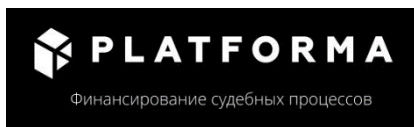
# Что в LegalTech является или становится AI/ML



Автоматические ответы на вопросы



Автоматические ответы на вопросы (нейросеть)



Финансирование судебных процессов, возможен рекомендательный поиск инвесторов и адвокатов



Документооборот на основе модели событий



Специализированные модели информационного поиска



\* По материалам конференции Skolkovo LegalTech, 1 декабря 2017

# Мозговой штурм

- Какую задачу в LegalTech Вы бы хотели решить с помощью ИИ?
- Какие для этого есть данные?
- Как организовать сбор необходимых данных?
- Как обеспечить качество данных?
- Нужна ли разметка или ассессорские оценки для данной задачи?
- Возможно ли сделать конкурс на открытых данных?
- Как организовать краудсорсинг?
- Гибридный интеллект:  
как встроить активное обучение в бизнес-процесс?