



Министерство экономического развития
Российской Федерации

РОССИЯ –
СТРАНА
ВОЗМОЖНОСТЕЙ

цифровой
прорыв 

сезон: ИИ

SibDS

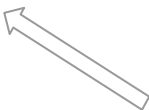
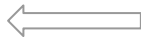
Юридический помощник для проверки НПА

Проблематика

Обработка
большого
количества данных



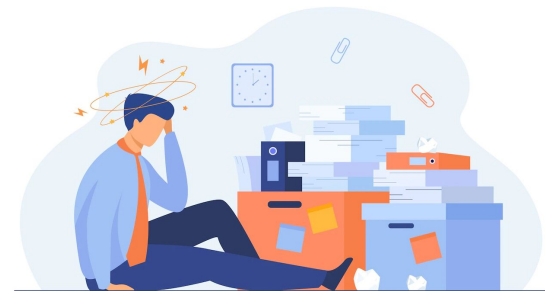
Сжатые сроки
для подготовки
проекта
документа



Проверка финальной
версии документа на
предмет логических,
смысловых и
текстовых ошибок



Сроки для
(межведомственного)
согласования проекта
документа





Использование ИИ в юриспруденции

Судебная система

Суммаризация и анализ
судебной практики

Подготовка проекта
судебного решения

Исследование
юридической литературы

Гражданское право

Подготовка примера
договора (template)

Подготовка примерного
содержания договора
(boilerplate)

Трудовое и административное право

Подготовка проекта
решения для простых
казусов

Задача

Разработать web-сервис,
позволяющий **автоматически**
проводить анализ НПА,
отражающий полноту
документа и его соответствие
требованиям.



Решение

1. Были **обучены** и **сравнены** между собой несколько **моделей нейронных сетей** и **машинного обучения**. **Итоговое решение** позволяет получить анализ с использованием **трех различных моделей**.
2. Был **разработан web-сервис** с удобным интерфейсом, позволяющий автоматически получать общий, а также более подробный **анализ НПА на предмет полноты и соответствия требованиям**.





Министерство экономического развития
Российской Федерации



цифровой прорыв 

сезон: ИИ

SibDS

Прототип



Техническая реализация

Парсинг документов

Извлечение текста из документов и разделение его на абзацы. Производилась **дополнительная сегментация текста**, отбирались необходимые для анализа разделы.



Классификация абзацев

Обучение и оценка моделей (LogReg, SVM, XGBoost, Ensemble - Stacking, BERT) на размеченных данных

Предсказание принадлежности абзаца к одному из классов (с определенной вероятностью)



“Объяснение” предсказаний моделей

Выделение ключевых слов, которые больше всего повлияли на отнесение абзаца к одному из классов



Новаторские идеи, фишки проекта

1. Возможность получить **отчет** сразу от **нескольких моделей**
2. Использование **ансамблевого подхода** (Stacking) в качестве одной из моделей - позволяет улучшить качество
3. **Дополнительная сегментация текста** документа для повышения точности модели
4. Наличие **развернутого отчета**, отражающего принадлежность абзацев к классам, а также **выделение ключевых слов**, на которые модель “обратила внимание” при классификации



Масштабируемость

Предложенное решение является **универсальным** и **адаптивным**:

- Представленный web-сервис можно использовать для **анализа различных документов** заданного формата;
- Есть возможность **дообучить модели** на данных из **новых областей** и **встроить эти модели в пайплайн**.



Оценка качества алгоритмов

model	accuracy
LogReg	67,3
SVM	68,1
Ensamble	68,5
BERT (max token=128)	68,4
BERT (max token = 256)	69,1



Дальнейшая реализация решения

Хакатон

- Анализ данных, исследование проблематики и существующих решений;
- Построение и исследование моделей, сравнение их по метрикам качества;
- Разработка web-сервиса, позволяющего получить отчет по загруженному НПА о его полноте и соответствии требованиям.

Ресерч

- Дополнительный сбор данных - размеченных вручную/ сгенерированных синтетически;
- Улучшение текущих моделей путем дообучения;
- Написание документации по работе с сервисом.

Пилотирование

- Тестирование и оценка работы сервиса юристами, предложение возможных улучшений интерфейса/ функционала.

Продакшн

- Внедрение итоговой системы в бизнес процессы заказчика.



Министерство экономического развития
Российской Федерации

РОССИЯ –
СТРАНА
ВОЗМОЖНОСТЕЙ

цифровой
прорыв 

сезон: ИИ

SibDS

Наша команда



Анна Власова
@anvlasova

Анализ данных



Андрей Власов
@vlasov_ds

Анализ данных,
Back-end



Егор Иванов
@egoriv

Front-end



Алексей Перминов

Юридическая
консультация