

цифровой
прорыв 

сезон: ИИ

Хакатон цифровой прорыв

12-14 апрель

Команда Dark Flow

Кейс:

Семантическая классификация
текстов



Министерство
экономического развития
Российской Федерации

РОССИЯ –
СТРАНА
ВОЗМОЖНОСТЕЙ



Команда Dark Flow



Некрасова Людмила
UX/UI designer,
frontend developer

Python, Figma,
Дизайн веб-интерфейсов

luda.nekrasova@gmail.com



Некрасова Татьяна
Data scientist,
backend developer

Python, ML/DL, Project Management

tigra1977@yandex.ru



Дворников Артур
Speaker,
product manager

Market research and analysis, product
development methodologies

dvornikov_artur550@mail.ru

Проблематика

Отсутствует автоматизированная система проверки содержания документов, что требует постоянного вмешательства сотрудников.

Сотрудники тратят значительное время на просмотр содержания поступивших документов и их соответствия необходимому перечню.

Неэффективная сортировка документов по содержанию и формату, особенно при большом объеме файлов.



Описание решения

Решение: Создание программного модуля для автоматической семантической классификации документов на основе текста.

Задачи

1. Семантическая Классификация

Автоматическое определение типа документа на основе анализа семантики текста.

2. Точность и Надежность

Применение алгоритмов машинного обучения обеспечивает высокую точность классификации документов и минимизирует вероятность ошибок

3. Автоматизация процесса

Автоматическая фильтрация и сортировка документов в соответствии с их семантическим содержанием

Уникальность

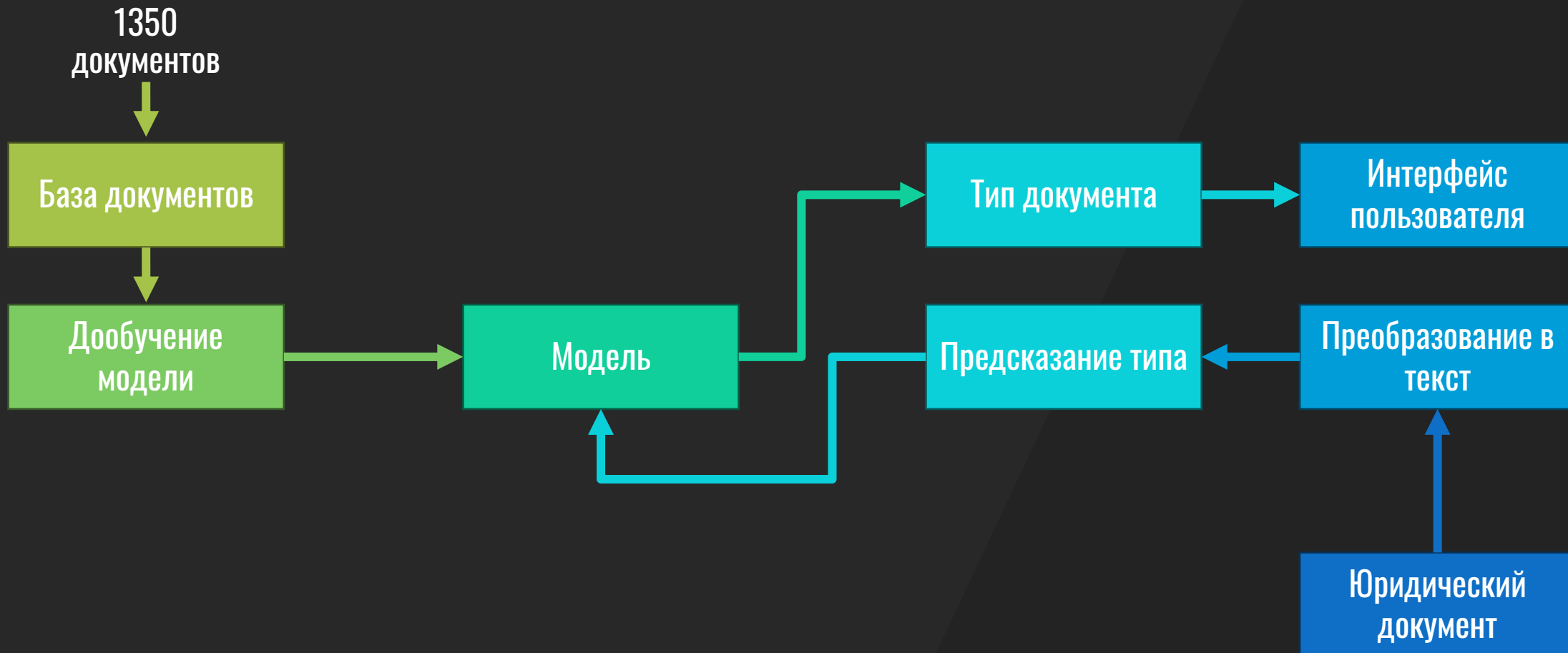
- ✓ Многоформатность
Поддержка различных форматов (PDF, DOCX, XLS).
- ✓ Предсказание по неполному тексту документа.
Анализ документов даже при частичном содержании.

Стек технологий



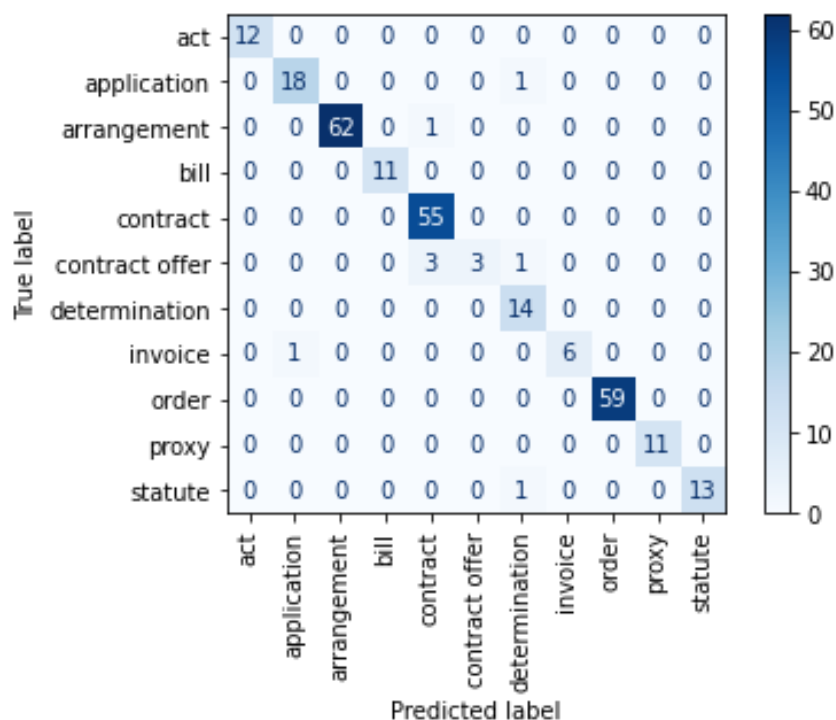
Архитектура

цифровой
прорыв



Предсказание типа документа

Высокая точность



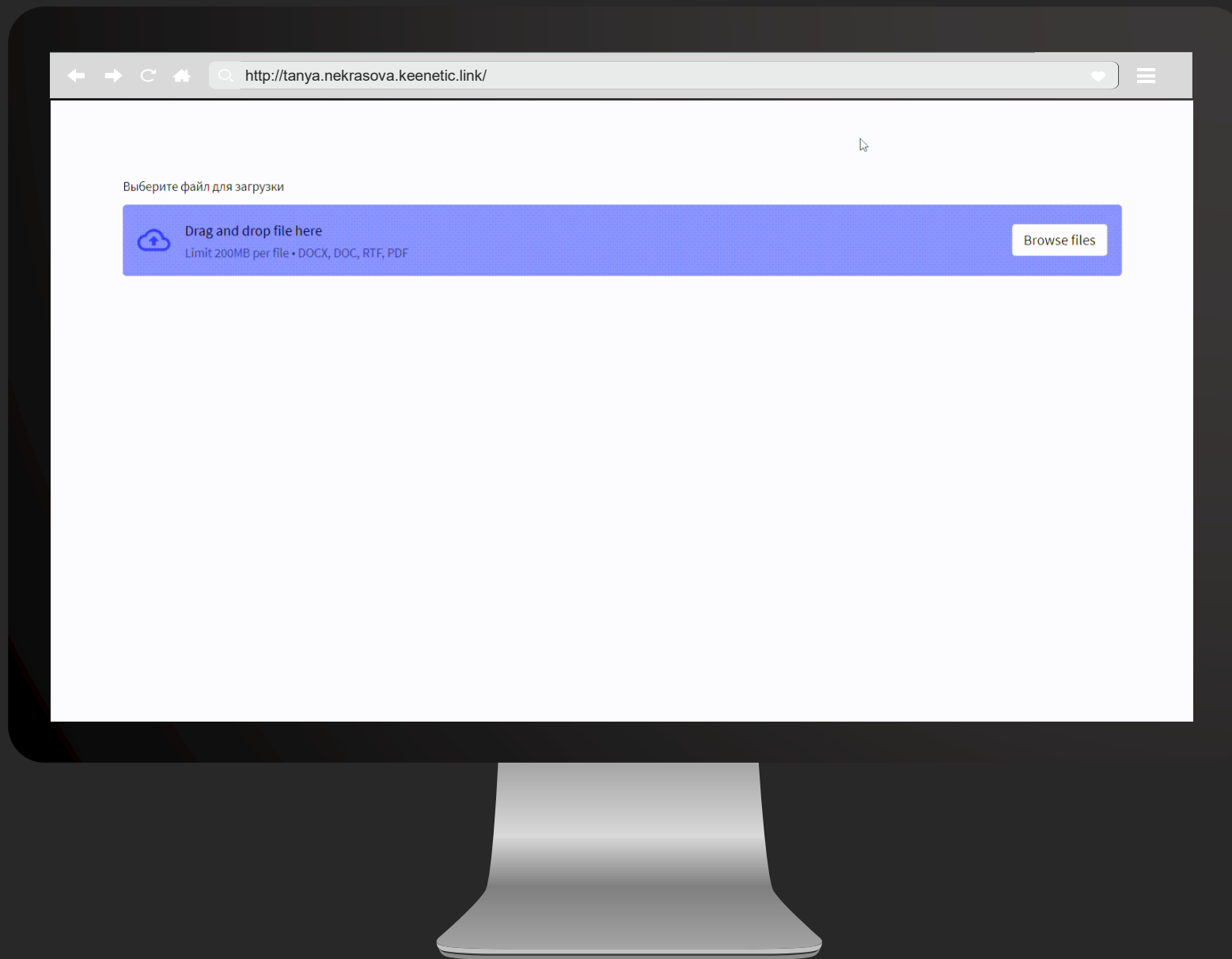
Тип документа	Обозначение
proxy	доверенность
contract	договор
act	акт
application	заявление
order	приказ
invoice	счет
bill	приложение
arrangement	соглашение
contract offer	договор оферты
statute	устав
determination	решение

Валидация
f1_weighted 0,9

Тест
f1_weighted 0,6

Демонстрация

цифровой
прорыв



Перспективы



Добавление новых типов
документов



Поиск по ключевым
словам



Расширение базы
данных
для улучшения обучения
моделей и оценки
производительности

Спасибо за внимание



Сайт

<http://tanya.nekrasova.keenetic.link/>