

MORE.Tech

4.0

TMteam

Профильная лента бизнес-новостей

Анализ задачи

Ролевая рекомендательная модель

- Фокус-группа с бухгалтерами для уточнения специфики потребления новостей
- Исследование тематических форумов и hh
- Выявление якорных особенностей ролей, формирование словаря ключевых интересов



Исходные данные

Источники данных RSS-ленты

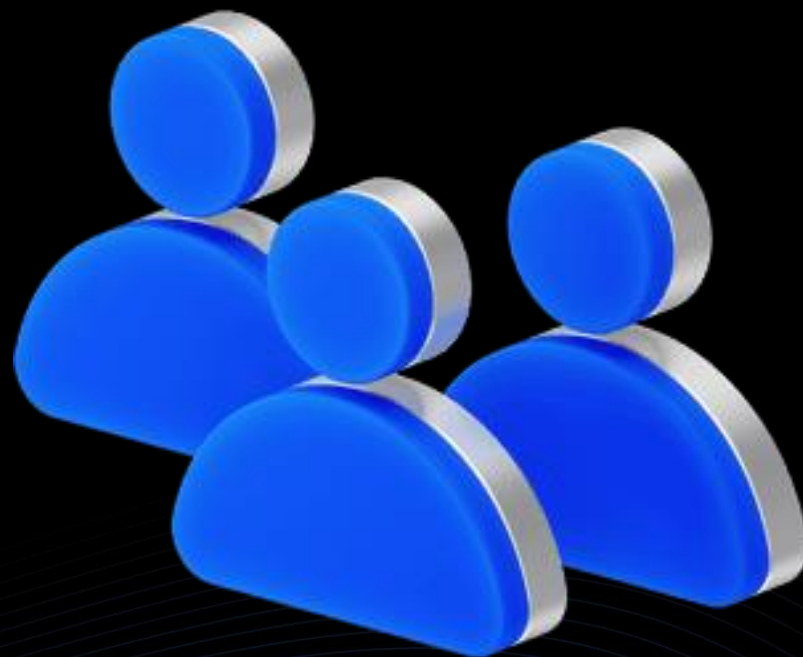
- СМИ
- Ведомства (ЦБ, министерства)
- Тематические подборки (Консультант)

Почему:

- Быстро
- Надёжно
- Масштабируемо

Источники данных для обучения модели

- Публичный архив новостей (Lenta.ru)



Общая часть

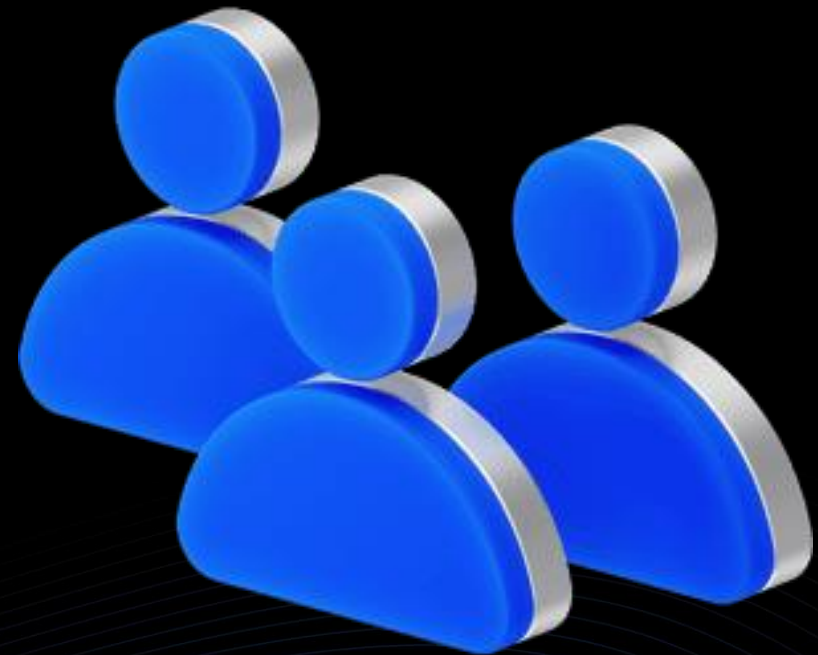
- Парсер RSS-лент
- Конфигуратор

Ветка 1 (алгоритмический прогноз)

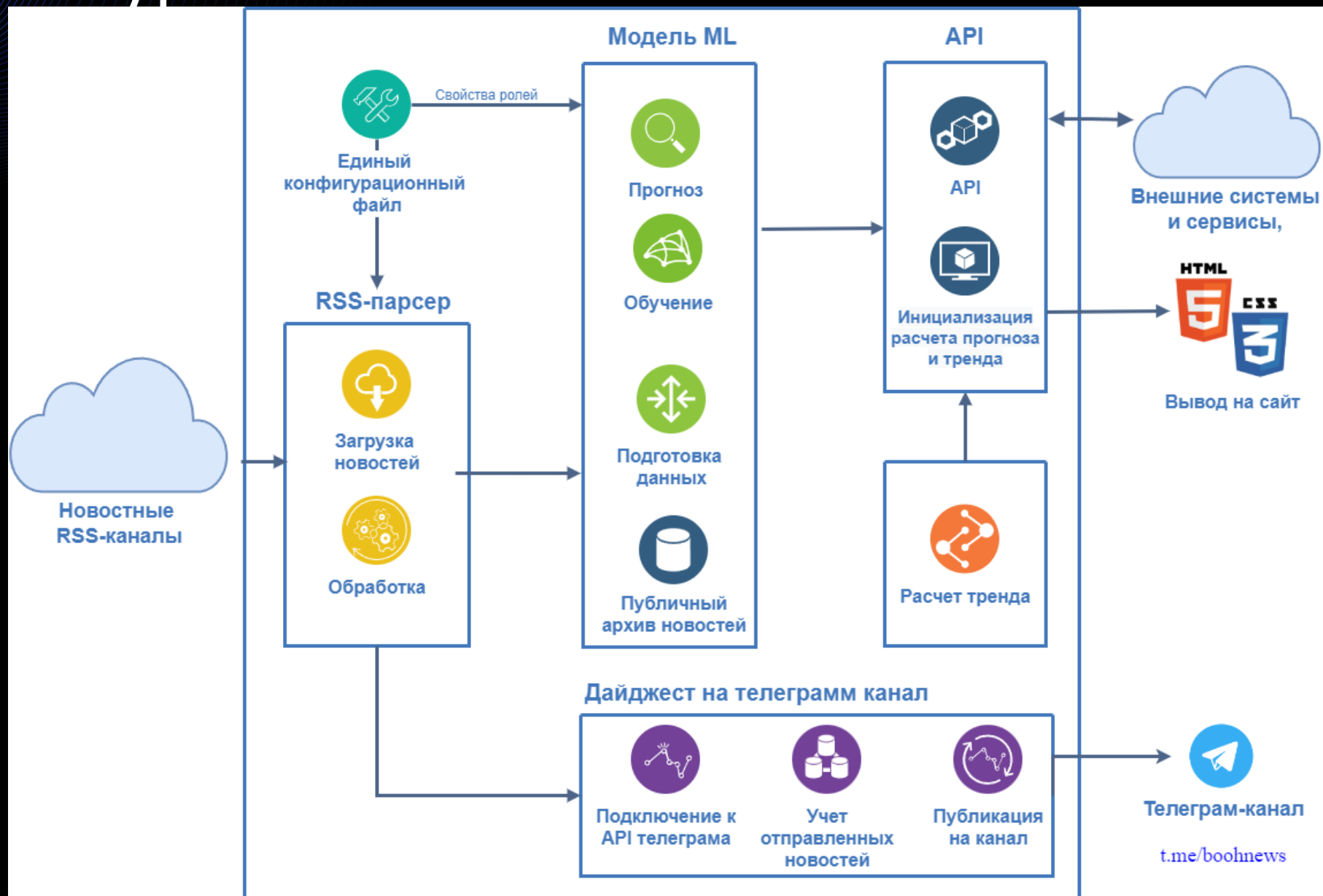
- Поиск новостей через алгоритмы схожести
- Поиск трендов: выделение сущностей, частотный анализ
- Публикация в телеграм-канале

Ветка 2 (ML прогноз)

- Обучение нейронной сети
- Реализация полноценного API
- Публикация на WEB-странице



Архитектура



Ветка 1

Алгоритмический прогноз

Тренды и инсайты

- Выделение сущностей из текста по схеме ROOT+NOUN+NER
- Составление матрицы распространения
- Расчет частот появления (для источника)
- Расчёт скорости и охвата распространения (между источниками)

Бабушка на лавочке рекомендует
4 подписчика

Организация представила в инспекцию справки с НДС. Налоговики обнаружили в них ошибки: не было ИНН налогоплательщиков, даты рождения указали как 01.01.1900. Организацию оштрафовали. Она обратилась в суд. Арбитры поддержали контролеров:
- любые сведения в справках...

http://www.consultant.ru/law/review/208422377.html#utm_campaign=rss_db&utm_source=rss_reader&utm_medium=rss 4 17:42

Бабушка на лавочке рекомендует
🔥🔥

Срок на возврат налоговой переплаты считают с даты подачи декларации, указала кассация

Организация обратилась к налоговикам за возвратом НДС. Контролеры отказали, так как налогоплательщик подал заявление несвоевременно. Спор дошел до суда. Первая инстанция и апелляция поддержали организацию: налогоплательщик подал заявление в пределах 3-летнего...

http://www.consultant.ru/law/review/208422378.html#utm_campaign=rss_db&utm_source=rss_reader&utm_medium=rss 4 17:42

9 октября

Вы подписались на канал

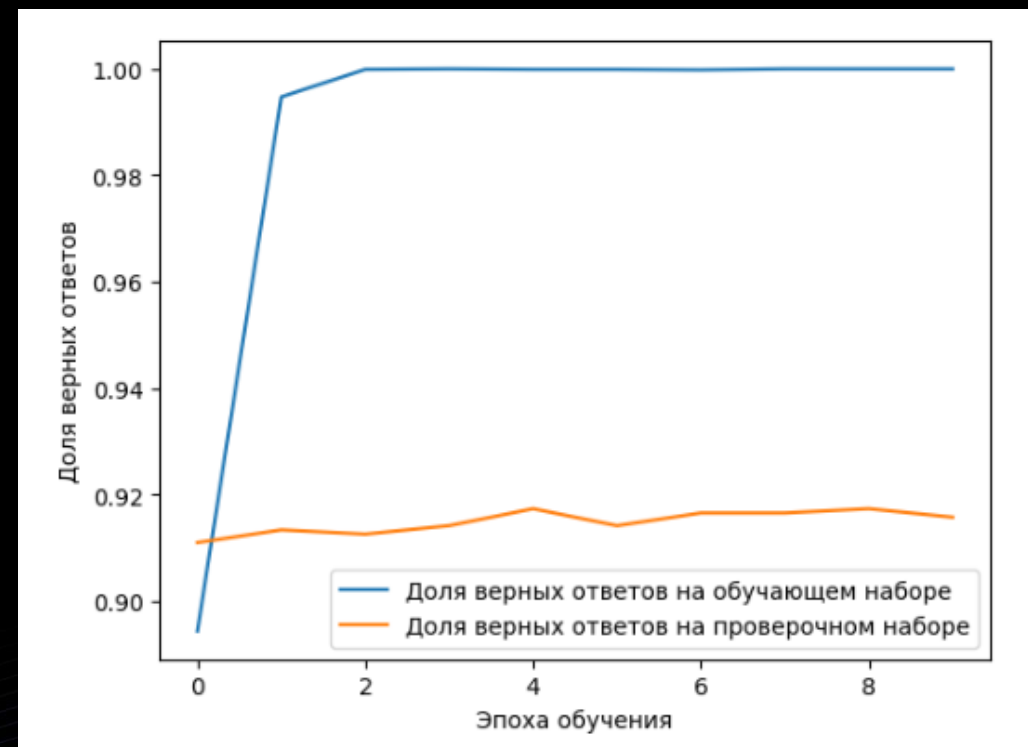
ОТКЛ. УВЕДОМЛЕНИЯ

Ветка 2

ML прогноз

Особенности решения

- Формирование обучающей выборки с учетом интересов конкретной роли
- Нейронная сеть, модель (BoW) бинарной классификации для под конкретную роль.
- тематические теги RSS осознано не использовались (только текст)



Особенности решения

- Инициализация сбора и обработки новостей, и прогнозирования
- Реализация полноценного API
- Публикация на WEB-странице



Подход в решении

Особенности

- Акцент на качестве кода (выделение всего в отдельные модули и классы, отказ от ноутбуков, переиспользование кода и пр.)
- Единый конфигурационный файл, через который, в том числе масштабируются роли (заведение дополнительных rss-каналов, изменение словаря ключевых слов роли)
- Упрощение задачи, отказ от сложных решений
- Регулярное обсуждение гипотез и идей



Команда



АЛЕКСАНДР

Капитан команды, аналитик,
разработчик



ЕКАТЕРИНА

Художник
Визуализация



ИВАН

Разработчик
ML-инженер



ЕВГЕНИЙ

Аналитик BIDS
Backend разработчик

Вопросы