Система рекомендации онлайн-курсов

Спикер: Тимофей Гриценко



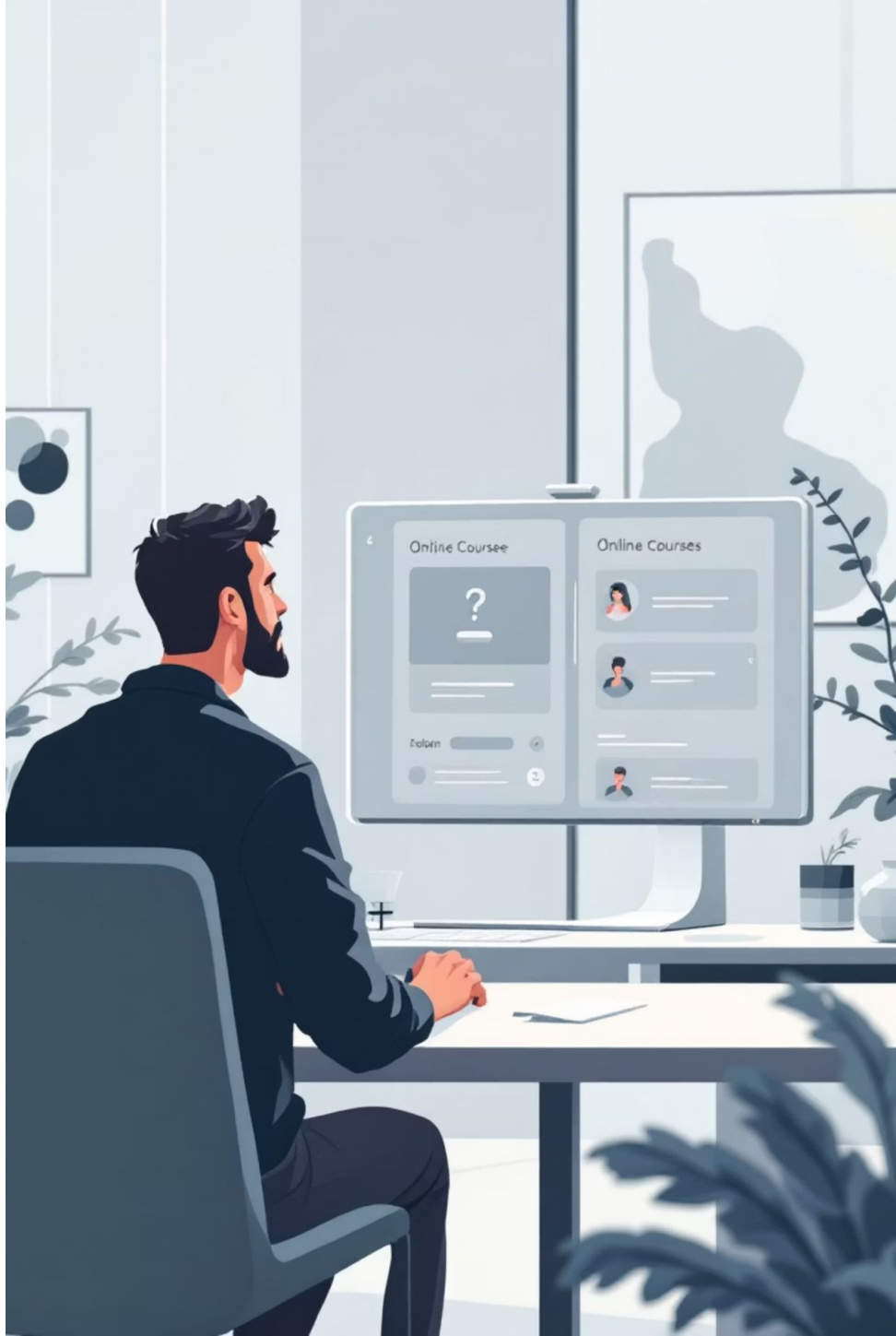
Проблема: Люди выбирают онлайн-курсы вслепую

Низкая завершаемость курсов

- Медианная доля завершения MOOC курсов составляет всего **12,6%**.
- Типичные исследования показывают, что только **7-13%** студентов доходят до финала.
- До **90%** пользователей онлайн-курсов (по данным MOOC) не завершают обучение.

Неоправданные ожидания

- Примерно **29%** опрошенных, бросивших курс, указали, что он *"не соответствовал ожиданиям"*, что говорит о неверном выборе.
- Хотя **64%** населения уже имели опыт онлайн-обучения и **58%** планировали учиться снова, эта проблема остается актуальной.



Решение проблемы: Интеллектуальный подход к выбору курсов

Реальные данные вместо звездочек

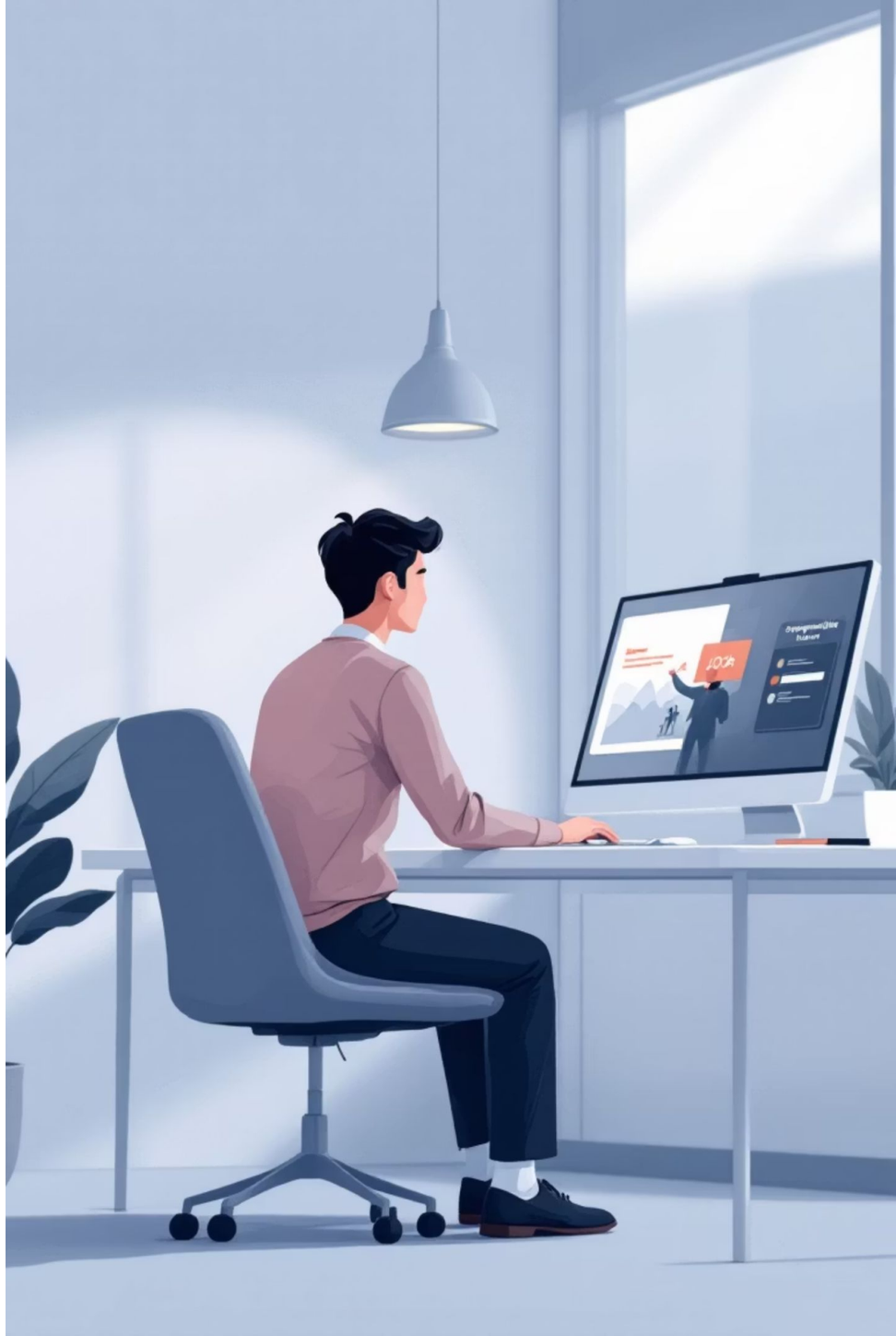
Мы оцениваем курсы на основе реальных карьерных результатов их выпускников.

Персональный подбор

Система собирает ваш профиль (уровень, бюджет, время, цель) и точно оценивает, насколько каждый курс подходит именно вам.

ML-ранжирование вместо маркетинга

XGBoost сортирует курсы по вероятности вашего успеха, а не по популярности или рекламным обещаниям.



Описание продукта: От хаоса к целенаправленности

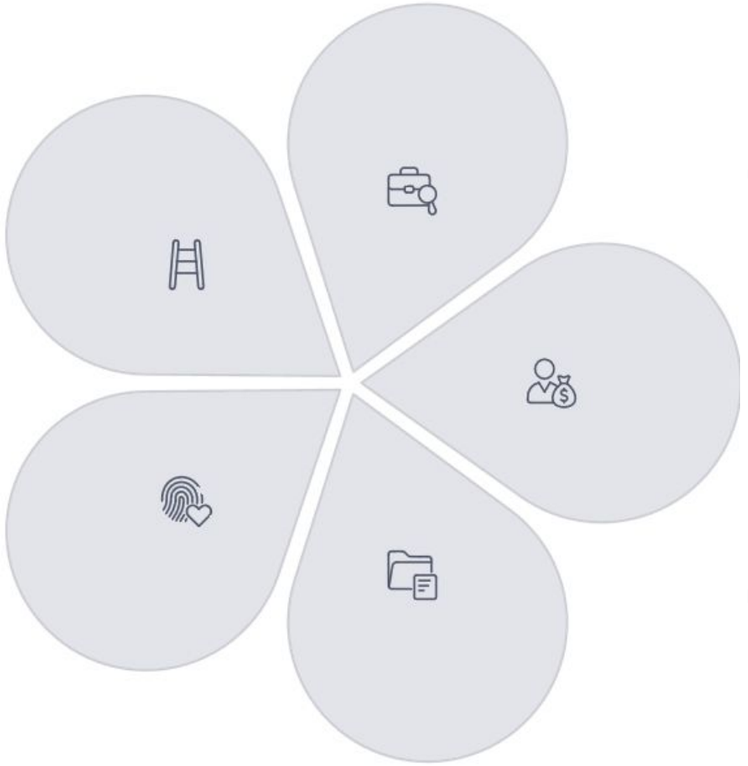
Наш продукт превращает выбор онлайн-курса из хаотичной и субъективной задачи в основанный на данных, персонализированный и карьерно-ориентированный процесс. Мы предлагаем пользователю курсы, которые с наибольшей вероятностью помогут ему достичь своей цели — перейти в новую профессию, получить работу или вырасти в доходе.

Карьерный рост

Помощь в продвижении по карьерной лестнице.

Персонализация

Индивидуальный подход к каждому пользователю.



Новая профессия

Облегчение перехода в новую область деятельности.

Увеличение дохода

Подбор курсов для повышения квалификации и зарплаты.

Данные в основе

Принятие решений на основе объективной информации.

Прогресс: Что уже сделано

Подготовка данных

Сгенерированы синтетические датасеты (курсы, студенты, вакансии, соц-сигналы, карьерная история) и реализована полная загрузка и очистка данных (нормализация дат, уровней, зарплат).

NLP-пайплайн

Построен NLP-пайплайн: извлечение навыков с помощью SkillNER + spaCy, анализ соц-постов, определение ролей и доменов.

Вычисление карьерных исходов

Реализована логика вычисления карьерных исходов для каждой пары (студент → курс).

Метрики эффективности

Построены агрегированные метрики эффективности курса (success rate, time-to-job, salary change и др.).

Профиль пользователя и ранжирование

Реализовано построение профиля пользователя и фич для ранжирования, а также добавлен ML-ранкер на XGBoost.

Интерактивный диалог (LLM)

Реализован диалог с пользователем на основе LLM: извлечение параметров профиля из свободного текста и генерация уточняющих вопросов.

Объяснения рекомендаций (LLM)

Добавлена генерация прозрачных объяснений выбора курсов на основе карьерных метрик и примеров выпускников.

Технологический стек: Основа нашего решения

Для создания нашей системы мы используем передовые инструменты и библиотеки, обеспечивающие высокую производительность и точность.

Python
(pandas, numpy) — обработка и анализ данных.

spaCy
NLP-обработка, парсинг текста и извлечение сущностей.

SkillNER
Специализированное извлечение навыков из текста.

Regex
Извлечение ролей, компаний и паттернов в данных.

XGBoost
(XGBRanker) — мощный инструмент для ML-ранжирования курсов.

pandas + datetime
Карьерные вычисления и агрегации временных рядов.

CSV-датасеты
Использование синтетических данных для прототипирования и тестирования.

Olamma
Используется для диалога с пользователем и генерации объяснений рекомендаций.



Возможные улучшения

Мы продолжаем развивать нашу систему, чтобы она стала еще более интуитивной и полезной для каждого пользователя. Впереди нас ждут новые этапы разработки.

Персонализированный образовательный путь

Построение не одного курса, а последовательности обучения

Анализ вакансии и резюме пользователя

Система анализирует текст вакансии и резюме пользователя, извлекает навыки и определяет недостающие компетенции

Прогноз карьерных исходов

Прогноз карьерных исходов и вероятности успеха для разных сценариев обучения

Улучшение пользовательского взаимодействия (UX / Interface)

Интерактивный веб-интерфейс с историей диалога, фильтрами по ограничениям и наглядным сравнением курсов по метрикам и навыкам



Команда: Кто стоит за проектом

Над созданием и развитием системы работают специалисты, объединенные общей целью — сделать онлайн-образование доступным и эффективным для всех.



Тимофей Гриценко

Machine Learning Engineer



Осипов Петр

Data Analyst



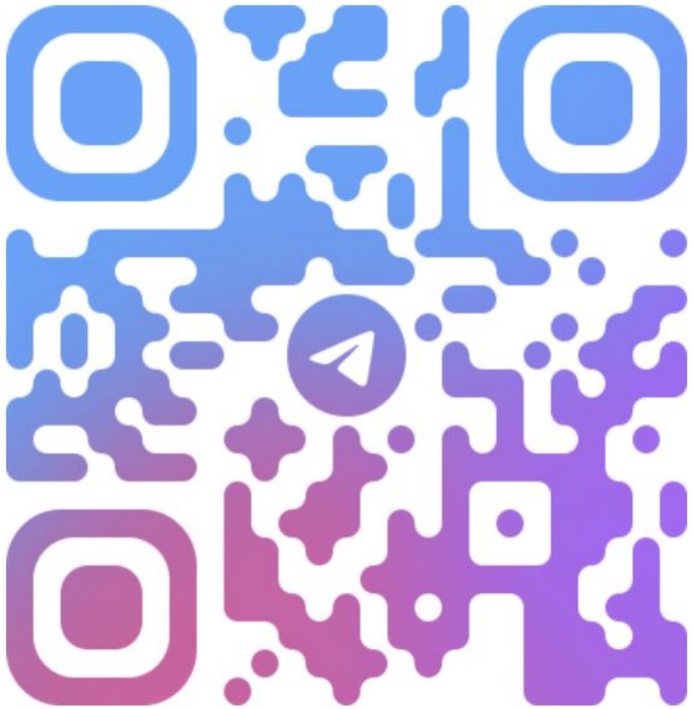
Лошманов Егор

Market Analyst

Контакты тимлида команды: Осипов Петр



почта: p.osipov1@g.nsu.ru



telegram: @O_osipovV