**Черновик решения**

от команды “Las Teteras Desesperadas”

**1. Введение**

Цель проекта

Реализовать самостоятельное ML-решение рекомендательной системы, способное сформировать персональные ранжированные списки направлений обучения в ВУЗах для каждого уникального запроса на основе данных, вводимых пользователем рекомендательной системы. Разработать документацию.

Основные задачи

* Разработка проектной документации (устава проекта, плана действий, таймлайна проекта, реестра рисков и черновика решения) и определение необходимых ресурсов.
* Сбор и обработка данных.
* Разработка ML-решения рекомендательной системы согласно спецификациям и требованиям проекта.
* Реализация ранжирования результатов по различным параметрам.
* Интеграция и тестирование.
* Разработка технической документации и документирование кода.

Более подробная информация представлена в уставе проекта (файл *Устав проекта.docx*).

**2. Описание решения**

Концепция IT-решения

Решение представляет собой разработку рекомендательной системы для облегчения выбора учебных заведений для студентов исходя из их предпочтений и интересов.

* Функциональность
  + Входные данные: информация пользователей сервиса о своих интересах, предпочтениях и активностях, включая фильмы, музыку, книги, игры, хобби и т. д., передаваемые по API сервиса.
  + Система рекомендаций: формирование на основе входных данных ранжированного списка рекомендаций по учебным заведениям и специальностям, наиболее соответствующих интересам и запросам пользователей.
* Важные особенности
  + ML-алгоритмы: использование современных алгоритмов машинного обучения для анализа данных пользователей и формирования рекомендаций на основе их предпочтений.
  + Ранжирование результатов по различным параметрам: предоставление возможности выдачи результатов по ВУЗам, факультетам и специальностям по отдельности, а также с ограничением по городу или региону, при указании в начале запроса до начала сравнения.
* Инновационные возможности:
  + Использование данных: эффективная обработка больших объёмов данных для предоставления более точных и персонализированных рекомендаций.
  + Гибкость и масштабируемость: создание архитектуры, позволяющей расширение функционала и лёгкую интеграцию с платформой заказчика.

Этапы реализации (план проекта)

* Инициация проекта
  + Встреча с заказчиком
  + Формирование проектной команды
  + Подготовка устава проекта
  + Определение необходимых ресурсов
* Планирование
  + Разработка плана действий
  + Составление таймлайна проекта и реестра рисков
  + Подготовка черновика решения
* Сбор и обработка данных
  + Сбор данных
  + Обработка данных
* Создание рекомендательной системы
  + Разработка ML-решения рекомендательной системы согласно спецификациям и требованиям проекта
  + Реализация ранжирования результатов по различным параметрам
* Интеграция и тестирование
  + Интеграция компонентов решения
  + Тестирование решения на совместимость и работоспособность
  + Доработка функционала
* Разработка документации
  + Создание технической документации
  + Документирование кода
* Завершение проекта
  + Подготовка отчёта о реализации проекта
  + Подведение итогов
* Дополнительные задачи (при наличии времени)
  + Реализация решения в виде Docker-контейнера
  + Реализация взаимодействия через API

Более подробная информация представлена в плане действий и распределении ресурсов (файл *План действий.docx*).

Планируемые к использованию технологии и инструменты

* язык программирования Python для написания кода проекта;
* библиотека Pandas для сбора и обработки данных;
* библиотеки Scikit-Learn, Natural Language Toolkit для построения модели;
* Google Colab для создания ноутбуков и проверки гипотез;
* Google Drive для хранения файлов и предоставления доступа к ним;
* сервис GitHub для создания репозитория и совместной работы над проектом.

Оценка рисков  
Список потенциальных рисков, связанных с реализацией проекта, и предложения по их управлению и минимизации:

* Неправильное планирование времени и ресурсов

Обоснование: Средняя вероятность неправильного планирования времени и ресурсов обусловлена сложностью оценки реальных затрат на разработку ML-решения. Высокое влияние говорит о том, что это может существенно замедлить процесс и снизить качество конечного продукта. Высокая величина риска требует регулярного мониторинга и корректировки действий. План действий включает разработку детального плана, а также регулярный контроль прогресса и адаптацию.

* Ошибки в работе алгоритмов рекомендаций

Обоснование: Низкая вероятность возникновения ошибок в работе алгоритмов обусловлена внимательной разработкой системы. Однако, высокое влияние указывает на то, что даже небольшие ошибки могут серьезно повлиять на результаты и точность рекомендаций. Средняя величина риска требует постоянного мониторинга работы алгоритмов и тестирования на различных данных.

* Проблемы с масштабированием и производительностью

Обоснование: Средняя вероятность возникновения проблем с масштабированием и производительностью связана с трудностями предвидения возможных сложностей при увеличении объема данных. Среднее влияние указывает на среднюю степень влияния этих проблем на успешность проекта. Средняя величина риска требует систематического тестирования системы на больших данных и оптимизации алгоритмов для повышения производительности.

* Изменения законодательства по обработке данных

Обоснование: Средняя вероятность изменений законодательства обусловлена изменчивостью нормативных актов. Высокое влияние указывает на значимость этих изменений для проекта. Средняя величина риска требует внимательного анализа изменений и своевременного информирования заказчика.

* Изменение требований заказчика

Обоснование: Низкая вероятность изменения требований заказчика говорит о стабильности запросов в ходе проекта. Однако, среднее влияние показывает, что такие изменения могут повлиять на ход работы. Низкая величина риска требует отслеживания триггеров и реакции при изменении требований.

* Болезнь или командировка участника команды

Обоснование: Низкая вероятность возникновения болезни или командировки участника указывает на стабильность команды. Среднее влияние указывает на среднюю значимость этого риска для успешного завершения проекта. Низкая величина риска требует подготовки замены участника команды для сохранения производительности команды в случае его отсутствия.

Более подробная информация представлена в реестре рисков (*файл Реестр рисков.xlsx*).

**3. Практическая ценность и применимость**

Улучшение образовательного процесса

Наше решение облегчит студентам процесс принятия решения при выборе образовательного учреждения и специальности, поможет им встретиться с людьми, в соответствии с их интересами и предпочтениями.

Практические преимущества использования

Наше решение предоставит персонализированные рекомендации на основе предпочтений в музыке, фильмах, книгах, играх, хобби, сократит время на поиск учебного заведения, увеличит вероятность выбора подходящей программы обучения.

Решение конкретных проблем и достижение поставленных целей проекта

* Сокращение времени принятия решения:
  + Проблема: Студентам часто требуется значительное время на поиск информации о университетах, специальностях и образовательных программах.
  + Решение: Рекомендательная система предоставит персонализированные рекомендации, исходя из интересов пользователя, что позволит сократить время, затрачиваемое на поиск подходящего университета и специальности. Пользователь получит отфильтрованный список рекомендаций, соответствующих его предпочтениям, что значительно упростит процесс принятия решения.
* Улучшение учебного опыта:
  + Проблема: Неподходящий выбор учебного заведения или специальности может негативно сказаться на учебном опыте студента.
  + Решение: Рекомендательная система предложит рекомендации, основанные на интересах студентов, что поможет выбрать программу обучения, соответствующую их предпочтениям. Это повысит вероятность удовлетворения студентами выбранными учебными заведениями и специальностями, что в свою очередь может положительно сказаться на их образовательном опыте и результативности обучения.
* Повышение вероятности выбора подходящей программы обучения:
  + Проблема: Множество ВУЗов, специальностей и программ обучения делают процесс выбора сложным и запутанным для студентов.
  + Решение: Рекомендательная система поможет студентам сузить выбор, предоставив ранжированные рекомендации на основе их интересов. Это уменьшит путаницу и позволит пользователям сосредоточиться на более ограниченном, но более релевантном списке ВУЗов и специальностей, снизив вероятность ошибочного выбора.
* Персонализированный подход к выбору образования:
  + Проблема: Традиционные подходы к выбору образования обычно не учитывают индивидуальные интересы и предпочтения студентов.
  + Решение: Рекомендательная система основывается на данных о предпочтениях пользователя, что позволит предложить персонализированные рекомендации. Этот индивидуальный подход поможет студентам обнаружить образовательные возможности, соответствующие их уникальным интересам и предпочтениям.

**4. Команда и план действий**

Члены команды

* Загальский Игорь - г. Апрелевка, заместитель начальника технического подразделения;
* Антропова Наталия - г. Москва, QA specialist портальных, web и enterprise решений, Sound producer;
* Майнгерт Владимир - г. Краснодар, инженер в сфере строительства аэродромов;
* Санникова Юлия - г. Краснодар, инженер-проектировщик (горный инженер) в частной компании;
* Мухачев Иван - г. Санкт-Петербург, логистика;
* Аванесян Тачат - Московская область, раньше работал в НИИ, последние 5 лет плавно переходит в аналитику.

Распределение ролей и ответственности

* проект-менеджер, дата-инженер - Загальский Игорь - составление документации проекта, сбор данных;
* продакт-менеджер, дата-аналитик, тестировщик - Антропова Наталия - разработка продукта, обработка и анализ данных, тестирование;
* дата-инженер - Майнгерт Владимир - сбор данных;
* ML-инженер - Санникова Юлия - разработка и обучение моделей;
* технические писатели - Мухачев Иван, Аванесян Тачат - разработка технической документации.

Каждый этап реализации проекта с оценкой времени, необходимого для выполнения каждой задачи членами команды, представлен в таймлайне проекта (файл *Таймлайн проекта.xlsx*)

**5. Заключение**

Основные достижения и преимущества нашего решения

Наше решение обладает несколькими ключевыми достижениями и преимуществами:

* Инновационные технологии и методы: Мы используем передовые технологии, включая алгоритмы машинного обучения и анализа данных, что обеспечит точность, гибкость и эффективность системы.
* Персонализация обучения: Наше решение обеспечит индивидуальный подход к образованию с учётом потребностей и интересов каждого учащегося, что поспособствует повышению мотивации и результативности обучения.

Практическая ценность и потенциал для улучшения образовательного процесса

Наше решение не только предоставит технологические новшества, но и привнесёт значительную практическую ценность в образовательный процесс:

* Индивидуализация обучения: Позволит студентам изучать материалы в соответствии с их уникальными потребностями и учебными темпами, создавая более эффективное образовательное окружение.
* Повышение доступности образования: Наше решение обеспечит гибкость в обучении, делая его доступным из любой точки мира, что поспособствует расширению образовательных возможностей для широкой аудитории студентов.
* Оптимизация образовательного процесса: Благодаря анализу данных и рекомендательным системам, наше решение поможет улучшить уровень обучения с учётом индивидуальных потребностей и предпочтений студентов.