ТЕХНИЧЕСКОЕ ЗАДАНИЕ (MVP)

Проект: Daryn Online
Интеллектуальная система прогнозирования академической успеваемости и раннего выявления образовательных рисков на основе искусственного интеллекта

Наименование	Модуль ИИ-прогнозирования успеваемости и раннего выявления рисков (MVP)
Компания-заявитель (заказчик)	ТОО «Daryn.online» — владелец и оператор образовательной платформы Daryn Online
Контактное лицо	kurmanbaevasymbat@gmail.com
Пользователи MVP	Преподаватели Daryn Online; администратор модуля
Данные	Открытые обезличенные образовательные датасеты из сети Интернет и синтетические данные, которые создаёт исполнитель. Реальные данные учащихся Daryn Online исполнителю не передаются
Срок разработки MVP	1 месяц (4 недели), этапы — в разделе 10
Дата	09.11.2026 (редакция 2)

1. Проблема

Снижение успеваемости учащегося сейчас становится заметно по факту — после выставления итоговых оценок, когда времени на исправление ситуации почти не остаётся. Преподаватель не успевает вручную отслеживать динамику результатов, пропуски заданий и падение активности каждого учащегося по всем классам, поэтому учащиеся группы риска выявляются поздно.

2. Цель MVP

За минимальный срок создать рабочий модуль, который по учебным данным прогнозирует итоговый балл учащегося, заранее — не менее чем за 3 недели до конца учебного периода — выявляет риск неуспеваемости и объясняет преподавателю его причины. MVP должен подтвердить, что модель выявляет риски точнее простого правила «средний балл ниже порога» и что её объяснения понятны и полезны преподавателю.

Модуль является самостоятельным продуктом: генерация тестов создаёт задания, адаптивное обучение подбирает траекторию учащемуся, а данный модуль анализирует данные и помогает преподавателю принимать решения. Модуль не изменяет учебный контент, решение о мерах всегда принимает преподаватель.

3. Границы MVP

Входит в MVP	Не входит (следующие версии)
Генератор тестовых данных и загрузка данных в CSV	Подключение реальных данных учащихся Daryn Online
Прогноз итогового балла по предмету	Прогноз результатов внешних экзаменов (ЕНТ)
Индекс риска (0–100) и уровень: низкий, средний, высокий	Типы риска и детекция аномалий в индивидуальной траектории
3 ключевых фактора-объяснения к каждому прогнозу	Рекомендации педагогических мер, учёт вмешательств и оценка их эффекта
Веб-панель преподавателя: список группы риска и карточка учащегося	Панели куратора и администрации, уведомления, родительский доступ
Экспорт списка группы риска в CSV	Отчёты PDF/XLSX
REST API модуля	SSO и встраивание в интерфейс Daryn Online
Язык интерфейса: русский	Казахская версия интерфейса

3.1. Учебные данные и источники данных

Источник данных в рабочей эксплуатации. Учебные данные (результаты тестов и заданий, сроки сдачи, журнал активности учащихся по классам и предметам) формируются на платформе Daryn Online, оператором которой является заказчик — ТОО «Daryn.online». Заказчик располагает этими данными и обеспечивает их защиту в соответствии с Законом Республики Казахстан «О персональных данных и их защите».

Данные в MVP. На всех этапах MVP используются два источника: открытые обезличенные образовательные датасеты из сети Интернет и синтетические данные, создаваемые исполнителем. Реальные данные учащихся и преподавателей Daryn Online исполнителю не передаются ни в исходном, ни в обезличенном виде и в модуле не обрабатываются. Заказчик предоставляет только описание структуры данных платформы (перечень сущностей и полей, типы значений) без самих записей.

Тестовые данные. Исполнитель разрабатывает генератор, который по описанию структуры формирует синтетические учебные данные в формате Daryn Online: условных учащихся, классы, предметы, результаты тестов и заданий, сроки сдачи и журнал активности. Объём — не менее 5 000 условных учащихся, 10 классов, 3 предметов и одного учебного периода (не менее 10 недель) с заранее известными итоговыми баллами. Данные строятся по сценариям: стабильная успеваемость, постепенное снижение, резкое падение, нерегулярная активность. Набор воспроизводим по фиксированному параметру (seed).

Контрольный набор. 20% условных учащихся откладываются и не используются при обучении; на них проверяются критерии приёмки (раздел 9).

Открытые датасеты. Исполнитель вправе использовать любые открытые образовательные датасеты из сети Интернет при соблюдении условий: данные обезличены, не содержат персональных данных, а лицензия допускает их использование в разработке. Примеры подходящих наборов: Open University Learning Analytics Dataset (OULAD) — результаты оценивания, итоговые результаты и журнал активности в системе обучения; UCI Student Performance — оценки по периодам и итоговые оценки учащихся школ; xAPI-Edu-Data — показатели активности и успеваемости учащихся. Перечень датасетов и их лицензии исполнитель фиксирует в отчёте о качестве.

Назначение открытых данных. Открытые датасеты используются для подбора признаков, обучения и дополнительной проверки моделей на реальных (обезличенных) образовательных данных, а также для настройки параметров генератора, чтобы синтетические данные были близки к реальным распределениям. Приведение открытых данных к формату Daryn Online выполняет исполнитель.

Не используются: персональные данные и результаты реальных учащихся, а также пол, национальность, социальный статус, состояние здоровья и иные чувствительные характеристики в качестве признаков модели.

3.2. Информационные системы и интеграции

Система	Назначение	Что требуется
Платформа Daryn Online (оператор — заказчик)	Источник учебных данных в рабочей эксплуатации; источник описания формата данных для MVP	Описание структуры данных (результаты тестов, задания, активность) без самих данных. Подключение реальных данных выполняет заказчик на своей стороне в следующей версии
Открытые образовательные датасеты (сеть Интернет)	Обучение и дополнительная проверка моделей, настройка генератора	Загрузка исполнителем из открытых источников; проверка лицензии и отсутствия персональных данных
Модуль прогнозирования (MVP)	Генерация синтетических данных, модели, API, веб-панель	Разрабатывается исполнителем
Хостинг (Alem Plus или иной)	Размещение модуля	Размещение инфраструктуры согласуется сторонами; GPU не требуется
БД и файловое хранилище	Открытые и синтетические данные, модели, прогнозы	PostgreSQL или SQLite в составе модуля

Обмен данными с модулем в MVP — загрузка CSV и REST API. Внешние LLM-сервисы в MVP не используются. Интеграция с банком заданий и интерфейсом Daryn Online в MVP не входит; модуль проектируется так, чтобы в следующей версии подключить реальный источник данных на стороне заказчика без переработки моделей и интерфейса.

3.3. Обязательства заказчика

ТОО «Daryn.online» подтверждает готовность:

- подтвердить, что является оператором платформы Daryn Online и располагает учебными данными; реальные данные исполнителю не передаются, для MVP предоставляется описание их структуры, а в качестве реальных обезличенных данных используются открытые датасеты;
- предоставить описание формата учебных данных платформы в течение первой недели;
- предоставить доступ к хостингу и репозиторию;
- назначить ответственного сотрудника для оперативных согласований;
- организовать экспертную апробацию с участием 5–10 преподавателей.

4. Пользовательский сценарий

1	Администратор генерирует тестовые данные или загружает CSV и запускает расчёт прогнозов.
2	Преподаватель открывает панель и выбирает класс и предмет.
3	Система показывает список учащихся с прогнозом итогового балла и уровнем риска; список можно сортировать и фильтровать.
4	Преподаватель открывает карточку учащегося: график динамики, прогноз и 3 фактора риска (например, «за 2 недели выполнено 2 из 6 заданий»).
5	Преподаватель выгружает список группы риска в CSV и сам принимает решение о мерах.

5. Функциональные требования

№	Функция	Требование MVP
1	Генерация тестовых данных	Не менее 5 000 условных учащихся по 4 сценариям; известные итоговые баллы; воспроизводимость по seed
2	Загрузка данных	CSV в формате Daryn Online; проверка структуры; сообщение об ошибке при неверном формате
3	Расчёт признаков	Окна 2 и 4 недели: средний балл, тренд, доля выполненных заданий, опоздания, активность
4	Прогноз балла	Прогноз итогового балла по предмету по 5-балльной шкале
5	Оценка риска	Индекс риска 0–100 и уровень: низкий, средний, высокий; пороги задаёт администратор
6	Объяснение	3 ключевых фактора к каждому прогнозу, сформулированные понятным преподавателю языком
7	Панель преподавателя	Список учащихся с риском, фильтр по классу и предмету, сортировка, карточка учащегося с графиком
8	Экспорт	Выгрузка списка группы риска в CSV
9	API	Список класса с риском; прогноз и факторы по учащемуся; запуск пересчёта; авторизация по ключу
10	Отчёт о качестве	Метрики модели на контрольном наборе и сравнение с правилом «средний балл ниже порога»

6. Контроль качества прогнозов

Перед показом прогноза преподавателю модуль выполняет автоматические проверки. Прогнозы, не прошедшие их, не показываются как надёжные и помечаются.

- **Достаточность данных:** при малом числе результатов у учащегося прогноз помечается «недостаточно данных».
- **Калибровка:** индекс риска соответствует фактической доле неуспевающих на контрольном наборе.
- **Согласованность:** прогноз балла и уровень риска не противоречат друг другу.
- **Объяснимость:** факторы рассчитаны для каждого прогноза и соответствуют данным учащегося.
- **Отсутствие перекоса:** качество модели проверяется отдельно по классам и предметам; разница Recall не более 10 п. п.

7. Технологическое решение

Компонент	Решение для MVP
Генерация данных	Python (NumPy, pandas); сценарии успеваемости с настраиваемыми параметрами
Модели	Градиентный бустинг (LightGBM или CatBoost): классификация риска и регрессия итогового балла; базовое правило для сравнения
Объяснимость	SHAP; перевод вклада признаков в текстовые формулировки по шаблонам
API	FastAPI (или согласованный), документация OpenAPI
Интерфейс	Веб-панель преподавателя (React, Vue или серверные шаблоны)
Хранение	PostgreSQL или SQLite: данные, модели, прогнозы
Развертывание	Docker, docker-compose; 2 vCPU, 4 ГБ RAM; GPU не требуется

Безопасность: вход по логину и паролю; две роли — преподаватель (только свои классы) и администратор; доступ к API по ключу; HTTPS; секреты хранятся вне кода; генератор не воспроизводит реальные ФИО, ИИН и контакты. Прогноз носит рекомендательный характер и не является оценкой учащегося.

8. Нефункциональные требования

- Расчёт прогнозов для 5 000 условных учащихся — не более 10 минут.
- Время ответа API и загрузки панели — не более 2 секунд.
- Одновременная работа не менее 10 преподавателей в MVP.
- Данные и модели не теряются при перезапуске контейнеров.
- Журнал действий: запуски генерации, обучения и пересчёта прогнозов.

9. Критерии приёмки MVP

Метрика	Целевое значение
ROC-AUC модели риска на контрольном наборе	≥ 0,75
Recall для уровня «высокий риск»	≥ 0,70
MAE прогноза итогового балла (5-балльная шкала)	≤ 0,6
Доля условных учащихся с итоговой неуспеваемостью, выявленных не менее чем за 3 недели	≥ 60%
Сравнение с правилом «средний балл ниже порога»	Модель точнее по ROC-AUC и Recall
Учащиеся сценария «стабильная успеваемость» с уровнем не выше «низкий»	≥ 90%
Оценка понятности и полезности объяснений преподавателями-экспертами	≥ 4 из 5
Реальные персональные данные в наборах данных	Отсутствуют

Проверка проводится на контрольном наборе тестовых данных; экспертная апробация — с участием 5–10 преподавателей заказчика на тестовых данных.

10. Этапы разработки

Этап	Срок	Результат
Подготовка и данные	Неделя 1	Согласование формата данных и сценариев, подбор открытых датасетов, генератор тестовых данных, расчёт признаков
Модели	Недели 2–3	Модель риска и прогноза балла, объяснения SHAP, отчёт о качестве со сравнением с базовым правилом
Панель и API	Недели 2–3 (параллельно)	Веб-панель преподавателя, API, развертывание в Docker
Испытания и апробация	Неделя 4	Проверка критериев приёмки, апробация с преподавателями, протокол, список доработок