

# Промежуточный отчет по программному проекту

## 1. Основные планы и этапы проекта

### 1.1 Краткое описание проекта:

«Приложение для анализа образовательных программ по Наукам о данных» представляет собой специализированный веб-сервис, предназначенный для сбора, обработки и анализа данных об образовательных программах в сфере наук о данных. Основной акцент сделан на автоматизированном сборе информации из различных источников, её структурировании в единую базу данных, применении инструментов машинного обучения для кластеризации и прогнозирования характеристик программ, а также на предоставлении визуальной аналитики через интерактивные графики и дашборды для облегчения сравнения и выбора подходящих вариантов.

### Название проекта:

«Приложение для анализа образовательных программ по Наукам о данных»

### Цель проекта:

Создать веб-приложение, которое автоматизирует процесс сбора и актуализации данных об образовательных программах по наукам о данных, обеспечивает их глубокий анализ с использованием методов машинного обучения, фильтрации и визуализации, способствуя обоснованному выбору для пользователей.

### Краткое описание задач:

1. Изучение предметной области, определение ключевых критериев для анализа программ по наукам о данных, обзор доступных источников данных и существующих аналогов.
2. Проектирование общей архитектуры системы и подготовка технического задания.
3. Сбор данных для приложения: парсинг информации, очистка от дубликатов, нормализация и загрузка в базу данных для дальнейшего анализа.
4. Разработка пользовательского интерфейса с поддержкой многоуровневых фильтров, группировок, визуализации данных и интеграцией ML-моделей.
5. Составление программной документации: техническое задание, пояснительная записка и исходный код программы.
6. Тестирование системы.
7. Защита проекта.

### 1.2 Планы и этапы выполнения проекта:

Этап проекта	Описание работ	Ожидаемые результаты	Сроки выполнения
--------------	----------------	----------------------	------------------

Анализ предметной области, постановка цели, анализ аналогов и источников данных	Изучение рынка образовательных программ, сбор литературы и определение ключевых метрик для анализа	Список атрибутов и метрик для данных	08.12.25 - 25.12.25
Разработка и утверждение технического задания	Определение требований к системе, согласование ТЗ с руководителем, загрузка в SmartLMS	Утвержденное ТЗ и план работ	09.12.25 - 16.12.25
Разработка программы	Проектирование структуры данных, API, прототипа UI (фильтры, дашборды, ML-интеграция)	Схема архитектуры, БД, прототип интерфейса	19.01.26 - 08.02.26
Подготовка и загрузка данных	Сбор данных из источников, очистка, валидация, нормализация, загрузка в БД	Заполненная БД, датасет для аналитики и ML	09.02.26 - 01.03.26
Аналитика и визуализация	Создание интерактивных графиков, расчет агрегатов, генерация выводов	Раздел аналитики с графиками	23.03.26 - 29.03.26
Элементы ML	Реализация кластеризации и рекомендаций на основе ML, интеграция в интерфейс	ML-модули и результаты в приложении	30.03.26 - 05.04.26
Разработка программной документации	Подготовка документации по ГОСТ 19 ЕСПД	Готовая документация	06.04.26 - 12.04.26
Испытания программы	Тестирование функционала, проверка ML-моделей и	Исправленные недочеты	13.04.26 - 19.04.26

	аналитики, исправление ошибок		
Подготовка и передача программы	Подготовка к презентации, защита, загрузка в ЛМС и Антиплагиат	Готовый продукт и защита	20.04.26 - 18.05.26

## 2. Используемый технологический стек и его обоснование

### 2.1 Перечень используемых технологий:

Технология/Инструмент	Описание	Причины выбора
Python	Язык для серверной логики, обработки данных и ML	Универсальный для аналитики, ML и бэкенда в едином стеке
PostgreSQL	База данных	Хранение структурированных данных, поддержка сложных запросов для аналитики
Pandas	Библиотека Python для предобработки и агрегации данных	Эффективная работа с табличными данными для аналитики
Plotly	Библиотека для интерактивных графиков	Обеспечивает визуализацию с интерактивностью

### 2.2 Обоснование выбранного технологического стека:

**Python:** Обеспечивает единую среду для ключевых задач — от сбора и обработки данных до аналитики и ML. Большое разнообразие библиотек ускоряет реализацию очистки, агрегаций и подготовки датасетов для визуализации и моделей.

**PostgreSQL:** Подходит для хранения карточек программ с атрибутами, выполнения фильтров и аналитических запросов (группировки по городам, стоимости), с поддержкой индексов для производительности.

**Pandas:** Используется для предобработки — нормализации данных, расчета средних значений, что важно для ML и визуализацией.

**Plotly:** Позволяет создавать интерактивные графики, ключевые для аналитики, с зумом и подсказками для лучшего понимания зависимостей.

**scikit-learn:** Закрывает задачи ML — кластеризацию по признакам и прогнозирование, с готовыми алгоритмами и стандартную последовательность этапов обработки данных (подготовка, обучение, применение модели).

### 3. Критерии оценивания проекта

Критерий	Описание
Качество сбора и обработки данных	Оценка полноты, точности и актуальности собранных данных, включая отсутствие дубликатов, нормализацию и валидацию датасета для аналитики и ML
Точность и эффективность ML-моделей	Процент точности кластеризации и прогнозирования (например, метрики Silhouette Score для кластеров, MAE/RMSE для регрессии), а также скорость вычислений моделей
Качество аналитики и визуализации	Наглядность графиков и дашбордов: наличие интерактивности, правильность агрегатов (средние, медианы), информативность текстовых выводов на основе данных
Покрытие тестами аналитических и ML-компонентов	Процент использования функциональности библиотек (Pandas, Scikit-learn, Plotly) для задач обработки, моделирования и визуализации данных
Эффективность использования стека для аналитики и ML	Интерактивность, легенды, подписи, тултипы
Качество интеграции данных с аналитикой	Скорость и корректность передачи данных из БД в ML-модели и визуализации, отсутствие ошибок при фильтрации и группировке