

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Программная инженерия»

СОГЛАСОВАНО

Доцент департамента программной
инженерии

_____ А. В. Меликян

«___» _____ 2026 г.

УТВЕРЖДЕНО

Академический руководитель
образовательной программы
"Программная инженерия",
старший преподаватель департамента
программной инженерии

_____ Н. А. Павловев

«___» _____ 2026 г.

**ПРИЛОЖЕНИЕ ДЛЯ АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ ПО
НАУКАМ О ДАННЫХ. ВЕБ-ПРИЛОЖЕНИЕ.**

Техническое задание

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.05.11-05 ТЗ 01-1-ЛУ

Исполнители:

Студент группы БПИ244

_____ / Н. Р. Яровицьна /

«___» _____ 2026 г.

2026

Инв.№ подп	Подп. и дата	Взам. инв.№	Инв.№ дубл.	Подп. и дата

УТВЕРЖДЕН

RU.17701729.05.11-05 ТЗ 01-1-ЛУ

**ПРИЛОЖЕНИЕ ДЛЯ АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ ПО
НАУКАМ О ДАННЫХ. ВЕБ-ПРИЛОЖЕНИЕ.**

Техническое задание

RU.17701729.05.11-05 ТЗ 01-1

Листов 28

Инв.№ подп	Подп. и дата	Взам. инв.№	Инв.№ дубл.	Подп. и дата

АННОТАЦИЯ

Техническое задание – это основной документ, оговаривающий набор требований и порядок создания программного продукта, в соответствии с которым производится разработка программы ее тестирование и приемка.

Настоящее Техническое задание на разработку «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение» содержит следующие разделы: «Введение», «Основания для разработки», «Назначение разработки», «Требования к программе», «Требования к программной документации», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки», приложения [7].

В разделе «Введение» указано наименование и краткая характеристика области применения программы.

В разделе «Основания для разработки» указан документ, на основании которого ведется разработка, и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение создаваемого программного продукта.

Раздел «Требования к программе» содержит указание на основные требования к функциональным характеристикам программы, к её надежности и к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению, а также специальные требования.

Раздел «Требования к программным документам» содержит указание на предварительный состав программной документации и специальные требования к ней.

Раздел «Технико-экономические показатели» содержит информацию об ориентировочной экономической эффективности разработки, экономические преимущества разработки программы.

Раздел «Стадии и этапы разработки» содержит информацию о стадиях разработки, этапах и содержании работ.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

Настоящий документ разработан в соответствии с требованиями:

1. ГОСТ 19.101-77 [1]: Виды программ и программных документов.
2. ГОСТ 19.102-77 [2]: Стадии разработки.
3. ГОСТ 19.103-77 [3]: Обозначения программ и программных документов.
4. ГОСТ 19.104-78 [4]: Основные надписи.

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

5. ГОСТ 19.105-78 [5]: Общие требования к программным документам.
6. ГОСТ 19.106-78 [6]: Требования к программным документам, выполненным печатным способом.
7. ГОСТ 19.201-78 [7]: Техническое задание. Требования к содержанию и оформлению.

Изменения к данному Техническому заданию оформляются согласно ГОСТ 19.603-78 [12], ГОСТ 19.604-78 [13].

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ	6
1.1. Наименование программы	6
1.2. Краткая характеристика области применения программы	6
2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ	7
2.1. Документ(ы), на основании которого(ых) ведется разработка	7
2.2. Наименование темы разработки	7
3. НАЗНАЧЕНИЕ РАЗРАБОТКИ	8
3.1. Функциональное назначение	8
3.2. Эксплуатационное назначение	8
4. ТРЕБОВАНИЯ К ПРОГРАММЕ	9
4.1. Требования к функциональным характеристикам	9
4.1.1. Требования к составу выполняемых функций	9
4.1.2. Требования к составу выполняемых функций	9
4.1.3. Требования к организации входных данных	11
4.1.4. Требования к организации выходных данных	13
4.1.5. Требования к временным характеристикам	15
4.1.6. Требования к интерфейсу	15
4.2. Требования к надежности	16
4.3. Условия эксплуатации	16
4.3.1. Климатические условия эксплуатации	16
4.3.2. Требования к видам обслуживания	16
4.3.3. Требования к численности и квалификации персонала	16
4.4. Требования к информационной и программной совместимости	16
4.4.1. Требования к серверному оборудованию	17
4.4.2. Требования к клиентскому оборудованию	17
4.5. Требования к информационной и программной совместимости	17
4.5.1. Требования к исходным кодам и языкам программирования	17
4.5.2. Требования к программным средствам, используемым программой	17
4.5.3. Требования к защите информации и программ	18
4.6. Требования к маркировке и упаковке	18
4.7. Требования к транспортированию и хранению	18

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4.8. Специальные требования	18
5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ	19
5.1. Состав программной документации	19
5.2. Специальные требования к программной документации	19
6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ	20
6.1. Предполагаемая потребность	20
6.2. Целевая аудитория	20
6.3. Преимущества перед аналогами	20
7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ	22
7.1. Стадии разработки, этапы и содержание работ	22
7.2. Сроки разработки и исполнители	23
8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ	24
8.1. Виды испытаний	24
8.2. Общие требования к приемке работы	24
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	25
ПРИЛОЖЕНИЕ. ССЫЛКИ НА АНАЛОГИ	27

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

1. ВВЕДЕНИЕ

1.1. Наименование программы

Наименование программы – «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение».

Наименование программы на английском языке – «Software for Analysis of Data Science Educational Programs. Web application».

1.2. Краткая характеристика области применения программы

В последние годы направление наук о данных (Data Science) приобретает всё большую популярность. Спрос на квалифицированных специалистов в этой области продолжает уверенно расти, что делает её одной из наиболее перспективных на рынке труда. Важным преимуществом профессии остаётся высокий уровень доходов.

Количество образовательных программ по наукам о данных стремительно увеличивается. Появляются новые предложения как в традиционном очном формате, так и в онлайн-режиме: от интенсивных курсов и до полноценных программ бакалавриата и магистратуры. Такой быстрый рост рынка образовательных услуг приводит к значительному разнообразию вариантов обучения. Абитуриенты сталкиваются с трудностями при выборе подходящей программы: сложно самостоятельно собрать и сравнить актуальную информацию о десятках предложений, оценить их по ключевым параметрам (стоимость, проходной балл, наличие бюджета, содержание и др.) и понять, какая из них лучше соответствует личным целям и возможностям.

В этих условиях возникает потребность в инструменте, который поможет систематизировать данные об образовательных программах по Data Science, поддерживать их в актуальном состоянии, предоставлять удобные средства фильтрации и сравнения, а также визуализировать статистику и проводить аналитику для более обоснованного выбора.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ

2.1. Документ(ы), на основании которого(ых) ведется разработка

Разработка ведется на основании учебного плана подготовки бакалавров по направлению 09.03.04 «Программная инженерия» и утвержденной академическим руководителем программы темы курсового проекта.

2.2. Наименование темы разработки

Наименование темы разработки - “Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение.”

Наименование темы разработки на английском языке - “Software for Analysis of Data Science Educational Programs. Web application.”

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

3. НАЗНАЧЕНИЕ РАЗРАБОТКИ

3.1. Функциональное назначение

Разрабатываемое приложение предназначено для анализа образовательных программ по наукам о данных. Продукт должен предоставлять следующие функции:

- Просмотр каталога образовательных программ;
- Многоуровневая фильтрация и группировка программ;
- Прямое сравнение выбранных программ;
- Построение интерактивных дашбордов;
- Формирование персонализированных рекомендаций;
- Кластеризация программ с использованием машинного обучения;
- Экспорт данных из каталога.

3.2. Эксплуатационное назначение

Приложение для анализа образовательных программ по наукам о данных – инструмент для анализа и сравнения образовательных программ в области наук о данных, анализа данных и машинного обучения с целью поддержки принятия решений при выборе образовательной траектории, а также для проведения аналитической оценки рынка образовательных услуг.

Основными конечными потребителями разрабатываемого приложения являются студенты, абитуриенты и специалисты, интересующиеся науками о данных, которые ищут удобный инструмент для анализа и выбора образовательных программ, а также аналитики образования, проводящие мониторинг и анализ программ по Data Science.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4. ТРЕБОВАНИЯ К ПРОГРАММЕ

4.1. Требования к функциональным характеристикам

4.1.1. Требования к составу выполняемых функций

4.1.2. Требования к составу выполняемых функций

Приложение должно обеспечивать выполнение следующих функций в области сбора, подготовки данных, аналитики и машинного обучения:

1. Сбор и управление данными

1.1. Автоматический сбор данных о образовательных программах по наукам о данных из открытых источников (например, сайты вроде Вузопедия, Учеба.ру и аналогичные).

а) Реализация периодического сбора данных (не реже одного раза в неделю);

б) Сбор ключевых атрибутов программ: название, факультет, уровень образования, вуз, город, стоимость, количество бюджетных и платных мест, проходной балл, длительность, форма обучения;

в) Обработка возможных ошибок сбора, включая логирование неудачных попыток и повторные запросы;

г) Обеспечение уникальности записей по комбинации полей (name, university, level) для предотвращения дубликатов.

1.2. Хранение собранных данных в реляционной базе данных с таблицей «programs» и следующими полями:

- id (INT) – уникальный идентификатор записи, генерируемый автоматически;
- program_code (TEXT) — код/направление из карточки программы в каталоге,
- name (TEXT) – название образовательной программы (например, «Прикладная математика и информатика»);
- faculty (TEXT) – название факультета (например, «Факультет компьютерных наук НИУ ВШЭ»);
- level (ENUM) – уровень образования; возможные значения: «бакалавриат», «специалитет», «магистратура»;
- university_name (TEXT) – полное название вуза (например, «Московский государственный университет имени М.В. Ломоносова»);
- city (TEXT) – город или регион вуза (например, «Москва», «Санкт-Петербург»);
- tuition_cost_rub_year (FLOAT) – стоимость обучения за год в рублях (NULL, если данные отсутствуют или программа полностью бюджетная);

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

- budget_places (INT) – количество бюджетных мест (0, если отсутствуют);
 - paid_places (INT) – количество платных мест (0, если отсутствуют);
 - budget_passing_score (FLOAT) – проходной балл на бюджет (NULL, если данные отсутствуют);
 - paid_min_score (FLOAT) – минимальный балл на платные места;
 - duration (INT) – длительность обучения в годах;
 - study_format (ENUM) – форма обучения: «очная», «заочная», «очно-заочная»;
 - language (TEXT) – язык обучения (например “Русский”);
 - accreditation (BOOL) — наличие аккредитации.

1.3. Валидация и очистка собранных данных:

- а) Проверка на полноту: обеспечение минимум 80% заполненности ключевых полей (name, level, university, city, passing score) для включения записи в базу;

б) Нормализация данных: приведение текстовых значений к стандартному формату (например, города к полным названиям: «Москва» вместо «мск» или «MSK»), стоимости к рублям без дополнительных символов; в) Обнаружение и обработка аномалий: удаление или корректировка записей с некорректными значениями (например, $\text{cost} < 0$, $\text{passing_score} > 300$ или < 0 , $\text{budget places} < 0$);

г) Логирование процесса очистки для отслеживания изменений.

1.4. Обновление датасета: поддержка ручного импорта данных из CSV-файлов администратором с автоматической валидацией формата и структуры (соответствие колонкам таблицы «programs»).

1.5. Экспорт данных: генерация полного или отфильтрованного датасета в форматах CSV или XLSX с заголовками на русском языке для дальнейшего анализа.

2. Аналитика и рекомендации с использованием машинного обучения

2.1. Кластеризация образовательных программ:

- а) Подготовка данных: выбор и извлечение до 5 признаков (стоимость, город, проходной балл, уровень образования, количество бюджетных мест и др.), с преобразованием категориальных признаков (например, кодирование для городов и уровней);

б) Настройка и выполнение алгоритма: поддержка алгоритмов кластеризации с параметрами (например, количество кластеров от 2 до 10);

в) Оценка качества: расчет метрик качества кластеризации;

г) Генерация описаний: автоматический расчет агрегированных статистик (средние, медианы, дисперсии и т.д.)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

анные, модальные значения) для каждого кластера и формирование текстового описания (например, «Кластер 1: Преимущественно магистратуры в Москве с высокой стоимостью (> 200 000 руб.) и проходным баллом > 80»);

д) Подготовка данных для визуализации: генерация координат для 2D-графика (с снижением размерности, если признаков > 2) с цветовой кодировкой кластеров.

2.2. Формирование персонализированных рекомендаций:

а) Сбор пользовательских предпочтений: обработка входных критериев (бюджет, предпочтительный город, уровень образования, желаемый проходной балл и др.) в виде вектора;

б) Применение алгоритмов: рекомендации на основе сходства между предпочтениями и программами;

в) Нормализация данных: унификация шкал (например, стоимость и балл);

г) Ранжирование и пояснения: формирование списка программ с процентом соответствия и текстовыми обоснованиями (например, «Соответствие 85%: Низкая стоимость (соответствует бюджету), город Москва (предпочтение)»);

д) Поддержка динамического обновления: пересчет рекомендаций при изменении предпочтений без полной перезагрузки данных.

2.3. Дополнительные функции машинного обучения:

а) Регрессионный анализ: предсказание проходного балла на основе соотношения бюджетных и платных мест, с подготовкой обучающего датасета и оценкой точности;

б) Классификация и сегментация: на основе атрибутов предсказывать категорию программы (например, высококонкурентная/стандартная/доступная). Пример: «Эта образовательная программа магистратуры классифицируется как высококонкурентная с вероятностью 85%».

4.1.3. Требования к организации входных данных

Входные данные для модулей сбора, подготовки данных, аналитики и машинного обучения включают пользовательские параметры, сырье данные из внешних источников, файлы для импорта и внутренние датасеты из базы данных.

1. Общие требования к обработке входных данных

1.1. Валидация входных данных: строгая проверка типов данных (например, cost как FLOAT, city как STRING), диапазонов значений (например, passing_score в пределах 0–300, budget_places ≥ 0) и форматов (например, отсутствие специальных символов в текстовых полях, кроме допустимых).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

1.2. Обработка ошибок: генерация понятных сообщений об ошибках на русском языке с указанием причины (например, «Некорректное значение стоимости: должно быть положительным числом»); возврат к предыдущему состоянию без потери данных.

1.3. Лимиты на входные данные: ограничение объема обрабатываемых записей (максимум 1000 записей для одного ML-запроса), количества признаков (до 5 для кластеризации по умолчанию) для предотвращения перегрузки сервера.

1.4. Простота ввода: входные данные не должны требовать от пользователя специальных знаний; использование интуитивных форматов (выбор из списков, диапазоны через интерфейс).

2. Входные данные для аналитики и визуализации

2.1. Параметры для дашбордов и графиков: тип графика (ENUM: линейный, столбчатый, круговой, точечный, гистограмма, картограмма), оси X/Y (из списка атрибутов, таких как cost, city, passing_score), группировка (ENUM: по city, level, form), агрегационные функции (ENUM: mean, median, count, min, max).

2.2. Передача параметров: в формате JSON через HTTP-запросы (POST для сложных настроек); серверная валидация на соответствие допустимым значениям и комбинациям.

2.3. Контекст запроса: учет текущего состояния сессии (примененные фильтры, выбранный датасет) для генерации актуальных данных.

3. Входные данные для машинного обучения

3.1. Параметры для кластеризации и рекомендаций: признаки (ARRAY строк из доступных атрибутов, без дубликатов), алгоритм (ENUM: K-Means, DBSCAN, Hierarchical), количество кластеров (INT в диапазоне 2–10), пользовательские предпочтения (DICT, например: {budget: FLOAT, city: STRING, level: ENUM}).

3.2. Подготовка данных: автоматическая нормализация числовых значений перед обработкой; проверка на полноту (минимум 80% заполненных полей в выбранных признаках).

3.3. Дополнительные параметры: для регрессии и классификации — целевой признак (STRING, например, passing_score), независимые признаки (ARRAY); ограничение на объем датасета для обучения.

4. Входные данные для сбора и управления данными

4.1. Источники для автоматического сбора: список URL открытых источников (в конфигурационном файле сервера); параметры парсинга (правила извлечения данных, такие как XPath или CSS-селекторы для каждого источника).

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4.2. Периодичность и логирование: входные параметры для запуска парсера (дата обновления, не реже еженедельно); обработка ошибок с повторными попытками и логированием.

4.3. Ручной импорт данных: файлы в формате CSV с обязательными полями (соответствующими схеме таблицы «programs»: name, level, university и т.д.); валидация на соответствие типов, полноту и отсутствие дубликатов перед вставкой в БД.

4.1.4. Требования к организации выходных данных

Выходные данные модулей сбора, подготовки данных, аналитики и машинного обучения включают результаты аналитики, визуализации, ML-моделей, экспортируемые файлы, обновленные датасеты и сообщения об ошибках.

1. Общие требования к выходным данным

1.1. Все выходные данные должны быть представлены на русском языке, с использованием понятной терминологии, без избыточных технических деталей (например, вместо «Silhouette Score: 0.5» — «Качество кластеризации: среднее»).

1.2. Формат: JSON для API-ответов с ключами, такими как «data», «visualization», «insights», «error»; файлы для экспорта (CSV, XLSX, PDF).

1.3. Выходные данные должны быть актуальными, основанными на текущем состоянии базы данных, с обязательным включением времени генерации.

1.4. В случае ошибок — стандартизованные сообщения на русском языке с описанием причины (например, «Ошибка в данных: некорректный формат входного файла»).

2. Выходные данные для визуальной аналитики и графиков

2.1. JSON-структура: {series: ARRAY (данные серий), labels: ARRAY (метки осей), insights: STRING (аналитические выводы), colors: ARRAY (цвета для элементов)}.

2.2. Текстовые выводы: краткие, информативные, размещаемые рядом с визуализацией (например, «Средняя стоимость в Москве выше на 20% по сравнению с другими городами»).

2.3. Экспорт: PDF-файлы с графиками, таблицами, водяными знаками даты и логотипом приложения; шрифт Arial, размер 12 pt; поддержка высокого качества изображений (PNG/SVG внутри PDF).

3. Выходные данные для машинного обучения и рекомендаций

3.1. Для кластеризации: JSON с {clusters: ARRAY {id: INT, description: STRING (автоматически сгенерированное текстовое описание)}, programs: ARRAY {program_id: INT, name: STRING}}, score: STRING (упрощенная оценка качества, например, «высокое» или «среднее»)}.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

3.2. Для рекомендаций: ранжированный ARRAY {program_id: INT, name: STRING, match_score: FLOAT (от 0 до 1, представленный как процент), explanation: STRING (краткое обоснование на русском)}.

3.3. Для дополнительных ML-функций (регрессия, классификация): JSON с {prediction: FLOAT или STRING (результат предсказания), metrics: DICT (упрощенные метрики, например, «точность: высокая»)}, explanation: STRING.

3.4. Обязательное предупреждение: «Рекомендации и предсказания основаны на алгоритмах машинного обучения – используйте как вспомогательный инструмент.»

4. Выходные данные для сбора и управления данными

4.1. Результаты парсинга: логи в формате TEXT или JSON с информацией о количестве собранных/обновленных записей, ошибках (например, «Собрано 150 программ, 10 ошибок парсинга»).

4.2. Обновленный датасет: автоматическое обновление БД с возвратом статуса (SUCCESS/ERROR) и статистики (количество новых записей, удаленных аномалий).

4.3. Ручной импорт: подтверждение в JSON {status: STRING, imported: INT, errors: ARRAY {message: STRING}}.

5. Экспорт данных

5.1. CSV: кодировка UTF-8, разделитель «;», заголовки на русском (например, «Название программы», «Уровень образования», «Стоимость» и т.д.); поддержка полного или отфильтрованного датасета.

5.2. XLSX: аналогично CSV, с форматированием таблиц (жирный шрифт для заголовков, автширина колонок).

5.3. PDF: с таблицами данных, графиками и выводами; шрифт Arial 12 pt, титульный лист с датой и названием отчета.

6. Требования к производительности и безопасности выходных данных

6.1. Анонимизация: отсутствие личных данных, конфиденциальной информации (например, внутренних ID или сырых запросов к БД).

6.2. Безопасность: выходные данные не должны раскрывать структуру БД или служебные детали; использование HTTPS для передачи.

6.3. Производительность: лимит на объем (например, максимум 10 000 строк в экспорте), сжатие файлов для больших датасетов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4.1.5. Требования к временным характеристикам

Требования к временным характеристикам программы не предъявляются.

4.1.6. Требования к интерфейсу

Интерфейс модулей аналитики, сбора данных и машинного обучения должен быть интуитивным, интегрированным с общим UI приложения, обеспечивать удобное взаимодействие пользователя с данными и инструментами ML, а также соответствовать общим требованиям к адаптивности, дизайну и навигации, описанным в основном техническом задании.

1. Общие требования к интерфейсу

1.1. Интерфейс должен быть полностью на русском языке с использованием профессиональной терминологии, без избыточных технических деталей для конечного пользователя.

1.2. Дизайн элементов: минималистичный, с высокой контрастностью, визуальной обратной связью (анимации при клике) и адаптивностью для десктопов, планшетов и мобильных устройств.

1.3. Интеграция с общей навигацией: доступ к разделам аналитики и ML через боковое меню (Графики и статистика, Умный анализ), с сохранением состояния сессии (примененные фильтры, выбранные данные).

2. Интерфейс визуальной аналитики и дашбордов

2.1. Конструктор дашбордов: поддержка перетаскивания виджетов для изменения положения, добавления/удаления элементов с автоматическим сохранением конфигурации.

2.2. Формы настройки: дропдауны для выбора типов графиков (столбчатый, круговой, линейный и др.), слайдеры для диапазонов значений (например, стоимость от X до Y), чекбоксы для группировки и агрегаций (mean, median, count).

2.3. Отображение результатов: интерактивные графики с зумом, подсказками при наведении; текстовые блоки под графиками с автоматически генерируемыми выводами (например, «Средний проходной балл в Москве: 85»).

2.4. Экспорт: кнопки для скачивания дашборда или отдельного графика в PDF/XLSX с пред просмотром.

3. Интерфейс машинного обучения и рекомендаций

3.1. Пошаговый мастер: последовательные шаги для настройки (выбор признаков через чекбоксы, алгоритма через дропдаун, параметров через слайдеры/инпуты), с кнопками «Назад» и «Далее» для удобства.

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

3.2. Интерактивные элементы: клик на кластер в графике показывает детали (список программ, описание); просмотр метрик качества при наведении.

3.3. Обработка процессов: индикатор прогресса (спиннер или бар) для длительных ML-вычислений; сообщения об ошибках на русском (например, «Превышен лимит данных, упростите запрос» или «Недостаточно данных для анализа»).

3.4. Страница рекомендаций: форма ввода предпочтений (поля для бюджета, города, уровня), отображение результатов в карточках с процентом соответствия и объяснением; кнопка «Обновить» для динамического пересчета.

4. Административный интерфейс для сбора и управления данными

4.1. Админ-панель: доступна только администраторам через отдельный раздел, с аутентификацией; кнопка «Обновить датасет» для запуска парсинга с подтверждением.

4.2. Логи и мониторинг: отображение лога ошибок в реальном времени (таблица с датой, описанием, источником); статистика сбора (количество новых записей, время выполнения).

4.3. Ручной импорт: форма загрузки CSV-файла с предпросмотром данных перед импортом, валидацией и сообщением о результате (успех/ошибки).

4.4. Экспорт: кнопки для полного или отфильтрованного датасета с выбором формата (CSV/XLSX).

4.2. Требования к надежности

Обработка ошибок ML, резервное копирование датасета ежедневно, мониторинг нагрузки на сервер.

4.3. Условия эксплуатации

4.3.1. Климатические условия эксплуатации

Специальные климатические условия не предъявляются.

4.3.2. Требования к видам обслуживания

Пользовательское обслуживание не требуется, так как работа осуществляется через браузер.

4.3.3. Требования к численности и квалификации персонала

Пользователь должен обладать навыками работы с веб-сайтами.

4.4. Требования к информационной и программной совместимости

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Приложение является веб-приложением клиент-серверной архитектуры и не требует специализированного оборудования. Развёртывание возможно на стандартных серверных платформах или облачных сервисах. Клиентская часть работает в веб-браузере пользователя.

4.4.1. Требования к серверному оборудованию

Рекомендуемые требования к серверному оборудованию для корректной работы приложения при любой нагрузке:

- 1) Процессор 4 виртуальных ядра или выше;
- 2) 8 ГБ оперативной памяти;
- 3) 40 ГБ SSD хранилища;
- 4) Поддержка контейнеризации (Docker) для запуска Python-приложений (FastAPI, ML-библиотеки).

4.4.2. Требования к клиентскому оборудованию

Клиентская часть представляет собой веб-интерфейс и работает в современном браузере. Специальных требований к оборудованию не предъявляется.

4.5. Требования к информационной и программной совместимости

Приложение должно использовать современный стек технологий, обеспечивающий высокую производительность, удобство разработки и лёгкость развёртывания.

4.5.1. Требования к исходным кодам и языкам программирования

4.5.1.1. Основной язык программирования — Python версии 3.10 или выше (для серверной части, сбора и обработки данных, аналитики и машинного обучения).

4.5.1.2. Исходный код должен соответствовать стандартам качества.

4.5.1.3. Обязательное использование системы контроля версий Git с удалённым репозиторием на GitHub или аналогичной платформе для совместной разработки.

4.5.1.4. Код должен быть оптимизирован для задач аналитики и ML: использование векторизованных операций в библиотеках, асинхронной обработки для парсинга данных и эффективного управления памятью при работе с большими датасетами.

4.5.2. Требования к программным средствам, используемым программой

Приложение должно использовать следующие программные средства и библиотеки:

4.5.2.1. Python — основной язык для серверной логики, сбора данных, обработки, аналитики и машинного обучения.

4.5.2.2. FastAPI — фреймворк для создания высокопроизводительного API с автоматической документацией и валидацией данных.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4.5.2.3. PostgreSQL — система управления реляционной базой данных для хранения структурированных данных о программах.

4.5.2.4. Pandas и NumPy — библиотеки для очистки, преобразования, агрегирования и анализа данных.

4.5.2.5. Plotly или Matplotlib — библиотеки для создания интерактивных веб-графиков и визуализации результатов аналитики.

4.5.2.6. Scikit-learn — библиотека для реализации задач машинного обучения, включая кластеризацию (K-Means, DBSCAN), рекомендации (KNN), регрессию и классификацию.

4.5.2.7. BeautifulSoup или Scrapy — библиотеки для веб-парсинга и автоматического сбора данных из открытых источников.

4.5.2.8. Docker и docker-compose — для контейнеризации и воспроизводимого развёртывания приложения, базы данных и ML-моделей.

4.5.2.9. Git + GitHub — для контроля версий и хранения исходного кода.

4.5.3. Требования к защите информации и программ

Требования к защите информации и программ не предъявляются.

4.6. Требования к маркировке и упаковке

Программа распространяется в виде электронного пакета, содержащего программную документацию, приложение и презентацию проекта. Специальных требований к маркировке и упаковке не предъявляется.

4.7. Требования к транспортированию и хранению

Транспортирование - посредством скачивания/клонирования репозитория;

Хранение продукта - в удаленном репозитории (GitHub) и локальной копии.

4.8. Специальные требования

Специальные требования к программе не предъявляются.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

5.1. Состав программной документации

1. «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение». Техническое задание (ГОСТ 19.201-78 [7]).
2. «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение». Пояснительная записка (ГОСТ 19.404-79 [10]).
3. «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение». Программа и методика испытаний (ГОСТ 19.301-79 [8]).
4. «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение». Текст программы (ГОСТ 19.401-78 [9]).
5. «Приложение для анализа образовательных программ по Наукам о данных. Веб-приложение». Руководство оператора (ГОСТ 19.505-79 [11]).

5.2. Специальные требования к программной документации

- 5.2.1. Документы к программе должны быть выполнены в соответствии с ГОСТ 19.106–78 [6] и ГОСТами к каждому виду документа (см. п. 5.1.).
- 5.2.2. Документация в формате .pdf и программа должны быть сданы в электронном виде в архиве формата .zip или .rar.
- 5.2.3. Документация должна быть загружена в дисциплину «Курсовой проект, ПИ 2 курс 25-26 уч. г.» в личном кабинете в информационно-образовательную среду Smart LMS «НИУ ВШЭ».

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

6.1. Предполагаемая потребность

Необходимость создания данной системы связана с увеличением интереса к образовательным программам в сфере Data Science, а также с потребностью в быстром и удобном сопоставлении предложений высших учебных заведений по ключевым формальным характеристикам.

6.2. Целевая аудитория

Пользователями системы являются абитуриенты и их родители, выбирающие образовательные программы по Data Science, а также студенты, сравнивающие варианты бакалавриата/магистратуры и других программ обучения. Кроме того, система может использоваться представителями образовательных организаций для анализа и сопоставления программ, а также специалистами и исследователями, занимающимися изучением рынка высшего образования.

6.3. Преимущества перед аналогами

Для проведения сравнительного анализа были выбраны ведущие российские агрегаторы образовательных программ («Postupi.online», «Табитуриент», «Вузопедия»).

Основное преимущество разрабатываемого приложения заключается в переходе от модели классического информационного справочника к интеллектуальной аналитической системе. В отличие от существующих аналогов, предоставляющих преимущественно статичные данные и базовую фильтрацию, предлагаемое решение реализует:

- Интеллектуальный анализ контента (ML): использование методов кластеризации для автоматической группировки программ по их реальному содержанию и семантической близости, а не только по формальным названиям.
- Прогностическую аналитику: внедрение предсказательных моделей.
- Глубокую визуальную аналитику: предоставление пользователю интерактивных дашбордов.
- Персонализированный подход: реализацию рекомендательной системы на основе машинного обучения, которая учитывает специфические требования пользователя.
- Профессиональный инструментарий: возможность экспорта структурированных аналитических данных (JSON, CSV, PDF) и наличие метрик оценки качества работы встроенных алгоритмов, что делает систему полезной не только для абитуриентов, но и для образовательных аналитиков и методистов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Таким образом, разрабатываемое приложение объединяет функции образовательного навигатора и инструмента Data Science, обеспечивая уникальный уровень прозрачности при выборе образовательной траектории.

Функция / Критерий	Мой проект	Postupi. online	Табиту- риент	Вузопе- дия
Автоматический сбор и обновление данных	+	+/-	-	+
Кластеризация программ с использованием ML	+	-	-	-
Персонализированные рекомендации на основе ML	+	+/-	+/-	+/-
Визуальная аналитика и дашборды	+	-	+/-	+/-
Предсказательные модели в аналитике	+	-	-	+/-
Оценка качества аналитики и ML	+	-	-	-
Экспорт и интеграция аналитических данных	+	-	-	+/-
Итого	7	1	1	3

Таблица 1. Сравнение функциональных характеристик

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

7.1. Стадии разработки, этапы и содержание работ

Стадии и этапы разработки были выявлены с учетом ГОСТ 19.102-77 [2].

Таблица 2 – Стадии и этапы разработки

Стадия разработки	Этап работ	Содержание работ	Сроки выполнения
Техническое задание	Научно-исследовательский этап разработки	Постановка задачи	26.11.2024
		Сбор исходных теоретических материалов	26.11.2024 – 10.12.2025
		Определение структуры входных и выходных данных	27.11.2024 – 15.12.2024
		Предварительный выбор методов решения задач	27.11.2024 – 15.12.2024
		Определение требований к техническим и программным средствам	27.11.2024 – 15.12.2024
	Разработка и утверждение технического задания	Обоснование возможности решения поставленной задачи	27.11.2024 – 15.12.2024
		Определение требований к программному продукту	27.11.2024 – 15.12.2024
		Выбор языков программирования	27.11.2024 – 15.12.2024
		Согласование технического задания и загрузка в SmartLMS	16.12.2024
	Сбор и подготовка данных	Поиск и парсинг источников данных о программах	03.02.2026 – 17.02.2026
		Очистка, нормализация и загрузка данных в БД	17.02.2026 – 03.03.2026
		Создание и наполнение базы данных PostgreSQL	03.03.2026 – 12.03.2026
		Проверка актуальности и полноты данных	12.02.2026 – 17.03.2026

Продолжение таблицы 2

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Стадия разработки	Этап работ	Содержание работ	Сроки выполнения
Рабочий проект	Сбор и подготовка данных	Документирование источников и структуры данных	19.03.2026
	Разработка frontend	Реализация визуальной аналитики и дашбордов (Plotly)	17.03.2026 – 31.02.2026
	Реализация ML-компонентов	Разработка модулей кластеризации и рекомендаций (scikit-learn)	20.02.2026 – 07.03.2026
		Интеграция ML в backend и тестирование	06.03.2026 – 16.03.2026
		Оптимизация производительности ML-запросов	16.03.2026 – 21.03.2026
	Финальная интеграция и отладка	Полная интеграция frontend, backend и ML	21.03.2026 – 31.03.2026
		Комплексная отладка и оптимизация производительности	31.03.2026 – 03.04.2026
		Разработка программной документации (ГОСТ 19 ЕСПД)	20.02.2026 – 03.04.2026
	Испытания программы	Разработка, согласование и утверждение методи	04.04.2026 – 08.04.2026
		Проведение испытаний и корректировка по результатам	08.04.2026 – 18.04.2026
	Внедрение	Подготовка к защите	18.04.2026 – 23.04.2026
		Защита	май 2026

7.2. Сроки разработки и исполнители

Разработка программного продукта должна быть завершена не позже утверждённого срока сдачи работы.

Исполнители – Яровицына Наталья Романовна.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ

Контроль и приемка разработки осуществляются в соответствии с документом «Программа и методика испытаний» (ГОСТ 19.301–79 [8]).

8.1. Виды испытаний

1. Полное и частичное функциональное тестирование.
2. Тестирование производительности.
3. Тестирование удобства пользования.
4. Тестирование безопасности.

8.2. Общие требования к приемке работы

Приемка работы может быть осуществлена только при корректной работе программы в соответствии с пунктом 4.1.1 при различных входных данных, соответствующих условиям в пункте 4.1.2 данного документа и при предоставлении полной документации к продукту, указанной в пункте 5.1, выполненной в соответствии с требованиями, указанными в пункте 5.2 данного технического задания.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. ГОСТ 19.101-77: Виды программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77: Стадии разработки. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77: Обозначения программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78: Основные надписи. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78: Общие требования к программным документам. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78: Требования к программным документам, выполненным печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78: Техническое задание. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.301-79: Программа и методика испытаний. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.401-78: Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 19.404-79: Пояснительная записка. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
11. ГОСТ 19.505-79: Руководство оператора. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. ГОСТ 19.603-78: Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
13. ГОСТ 19.604-78: Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
14. Официальная документация FastAPI [Электронный ресурс] / FastAPI. – Режим доступа: <https://fastapi.tiangolo.com/>, свободный. (дата обращения: 14.12.2025).
15. Официальная документация PostgreSQL [Электронный ресурс] / PostgreSQL. – Режим доступа: <https://www.postgresql.org/docs/>, свободный. (дата обращения: 14.12.2025).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

16. Официальная документация Plotly для Python (Graphing Library) [Электронный ресурс] / Plotly.
– Режим доступа: <https://plotly.com/python/>, свободный. (дата обращения: 14.12.2025).
17. Официальная документация Pandas [Электронный ресурс] / Pandas. – Режим доступа: <https://pandas.pydata.org/docs/>, свободный. (дата обращения: 14.12.2025).
18. Официальная документация scikit-learn [Электронный ресурс] / scikit-learn. – Режим доступа: <https://scikit-learn.org/stable/>, свободный. (дата обращения: 14.12.2025).
19. Официальная документация Docker [Электронный ресурс] / Docker. – Режим доступа: <https://docs.docker.com/>, свободный. (дата обращения: 14.12.2025).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ. ССЫЛКИ НА АНАЛОГИ

Приложение	Ссылка
Postupi.online	https://postupi.online/
Табитуриент	https://tabituirient.ru
Вузопедия	https://vuzopedia.ru

Дата обращения: 14.12.25.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-05 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ