

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук  
Департамент программной инженерии

**СОГЛАСОВАНО**

Доцент департамента математики  
факультета экономических наук, кандидат  
физико-математических наук

\_\_\_\_\_ Е. Р. Горяинова  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

**УТВЕРЖДАЮ**

Академический руководитель  
образовательной программы  
«Программная инженерия»  
профессор департамента программной  
инженерии, канд. техн. наук

\_\_\_\_\_ Н. А. Павлочев  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

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

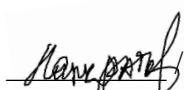
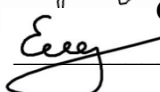
**Разработка программного комплекса для исследования  
влияния аномальных наблюдений на точность  
прогнозирования в регрессионных моделях**

Техническое задание

**ЛИСТ УТВЕРЖДЕНИЯ**

RU.17701729.11.04-01 ТЗ 03-1

**Исполнители:**

 Студент БПИ-221  
Панкратов С. Ю. /  
 Студент БПИ-221  
Знатнов Е. П. /  
«12» мая 2025 г

Москва 2025

УТВЕРЖДЕН

RU.17701729.11.04-01 ТЗ 03-1

**Разработка программного комплекса для исследования  
влияния аномальных наблюдений на точность  
прогнозирования в регрессионных моделях**

**Техническое задание**

**RU.17701729.11.04-01 ТЗ 03-1**

**Листов 27**

<i>Подп. и дата</i>	
<i>Инв. № дубл.</i>	
<i>Взам. инв. №</i>	
<i>Подп. и дата</i>	
<i>Инв. № подл</i>	

**Москва 2025**

## АННОТАЦИЯ

Техническое задание – 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 программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях содержит следующие разделы: «Введение», «Основание для разработки», «Назначение разработки», «Требования к программе», «Требования к программным документам», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки» и приложения.

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

В разделе «Основания для разработки» указаны документы, на основании которых ведется разработка, а также наименование темы разработки.

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

Раздел «Требования к программе» содержит основные требования к функциональным характеристикам, надежности, условиям эксплуатации, составу и параметрам технических средств, информационной и программной совместимости, маркировке и упаковке, транспортированию и хранению.

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

Раздел «Технико-экономические показатели» описывает ориентировочную экономическую эффективность, предполагаемую годовую потребность, а также экономические преимущества разработки по сравнению с аналогами.

Раздел «Стадии и этапы разработки» содержит стадии и этапы разработки, их содержание и сроки, а также указывает лица, ответственные за их выполнение.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы, а также зафиксированы все допустимые при этом виды испытаний.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Настоящий документ разработан в соответствии с требованиями:

1. ГОСТ 19.101-77<sup>[1]</sup>: Виды программ и программных документов.
2. ГОСТ 19.102-77<sup>[2]</sup>: Стадии разработки.
3. ГОСТ 19.103-77<sup>[3]</sup>: Обозначения программ и программных документов.
4. ГОСТ 19.104-78<sup>[4]</sup>: Основные надписи.
5. ГОСТ 19.105-78<sup>[5]</sup>: Общие требования к программным документам.
6. ГОСТ 19.106-78<sup>[6]</sup>: Требования к программным документам, выполненным печатным способом.
7. ГОСТ 19.201-78<sup>[7]</sup>: Техническое задание. Требования к содержанию и оформлению.
8. ГОСТ 19.602-78<sup>[8]</sup>: Правила дублирования, учета и хранения программных документов, выполненных печатным способом.

Изменения к настоящему техническому заданию должны быть оформлены согласно ГОСТ — 19.603-78<sup>[9]</sup> — и ГОСТ 19.604-78<sup>[10]</sup>.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

**СОДЕРЖАНИЕ****1. ВВЕДЕНИЕ5**

1.1. Наименование программы5

1.2. Краткая характеристика области применения5

**2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ6**

2.1. Документ(ы), на основании которого(ых) ведётся разработка6

2.2 Наименование темы разработки6

**3. НАЗНАЧЕНИЕ РАЗРАБОТКИ7**

3.1. Функциональное назначение7

3.2. Эксплуатационное назначение7

**4. ТРЕБОВАНИЯ К ПРОГРАММЕ7**

4.1. Требования к функциональным характеристикам8

4.1.2. Требования к организации входных данных12

4.1.3 Требования к организации выходных данных12

4.1.4. Требования к интерфейсу12

4.2. Требования к надежности13

4.3. Условия эксплуатации15

4.4. Требования к составу и параметрам технических средств15

4.5. Требования к информационной и программной совместимости16

4.6. Требования к маркировке и упаковке16

4.7. Требования к транспортировке16

**5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ17**

5.1. Предварительный состав программной документации17

5.2. Специальные требования к программной документации17

**6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ18**

6.1 Ориентировочная экономическая эффективность18

6.2 Предполагаемая потребность18

6.3 Экономические преимущества разработки по сравнению с отечественными и зарубежными образцами или аналогами18

**7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ19**

7.1. Стадии разработки, этапы и содержание работ19

7.2. Сроки разработки и исполнители21

**8. ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ22****СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.23****ТЕРМИНЫ И СОКРАЩЕНИЯ26**

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 03-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 1. ВВЕДЕНИЕ

### 1.1. Наименование программы

Наименование темы разработки: “Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях”.

Наименование темы разработки на английском языке: “Development of a Software Package to Study the Influence of Outliers on the Prediction Accuracy in Regression Models”.

Краткое наименование – “MSnOutliers”.

### 1.2. Краткая характеристика области применения

«MSnOutliers» – приложение для исследования качества различных статистических методов на выборках данных с большим числом зашумленных (т. е. содержащих в себе помимо полезной нагрузки некоторый шум известного распределения) данных.

Комплекс также интегрирует различные алгоритмы машинного обучения для обнаружения и устранения аномальных наблюдений, что позволяет исследовать их эффективность в повышении точности прогнозирования регрессионных моделей.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 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) ведётся разработка

Основанием для разработки является учебный план подготовки бакалавров по направлению 09.03.04 «Программная инженерия» и утвержденная академическим руководителем тема курсового проекта.

### 2.2 Наименование темы разработки

Наименование темы разработки: “Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях”.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

### 3. НАЗНАЧЕНИЕ РАЗРАБОТКИ

#### 3.1. Функциональное назначение

Приложение «MsnOutliers» предоставляет возможность изучить влияние аномальных наблюдений (данных, отличающихся по абсолютному значению на порядок от среднего по всей остальной выборке) на качество работы различных регрессионных методов математической статистики. Предполагается, что основное использование приложения будет происходить в образовательных целях.

#### 3.2. Эксплуатационное назначение

Приложение «MsnOutliers» позволяет отслеживать влияние выбросов в данных на качество работы регрессионных методов. Для анализа могут использоваться данные, подготовленные пользователем либо же сгенерированные на стороне приложения.

Целевая аудитория – школьники и студенты, проходящие подготовку по дисциплине “Математическая статистика” или каким-либо смежным с ней, а также преподаватели, читающие вышеупомянутые курсы.

Программный комплекс может использоваться на ПК с операционной системой Windows, Linux или MacOS. Предполагается, что основная часть комплекса будет написана на языке C++ с возможным использованием модулей на языке Python для визуализации разработанных в рамках проекта средств борьбы с влиянием выбросов и методов регрессии.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



## ТРЕБОВАНИЯ К ПРОГРАММЕ

### 4.1. Требования к функциональным характеристикам

#### 4.1.1. Состав выполняемых функций

##### 4.1.1.1. Работа с данными

**4.1.1.1.1** При запуске комплекса пользователь попадает на главный экран, где ему доступны основные функции для работы с данными:

- 1) Создание моделей регрессии
- 2) Сохранение конфигураций моделей
- 3) Запуск анализа
- 4) Генерация синтетических данных
- 5) Статистики целевой переменной

**4.1.1.1.2** Пользователь может загрузить свои данные в формате CSV при создании модели. CSV-файлы должны использовать точку с запятой (“;”) в качестве разделителя. Данные должны быть структурированы таким образом, чтобы признаки располагались в первых N столбцах, а целевая переменная - в последнем столбце. При создании модели пользователь указывает путь к файлу данных в поле “Path to data”, а также количество признаков в поле “Num. features”. Если файл не существует или имеет некорректный формат, пользователь получит соответствующее уведомление.

**4.1.1.1.3** Если пользователь не имеет своих данных для анализа, комплекс предоставляет возможность генерации синтетических данных. Для этого необходимо нажать кнопку “Generate data”, после чего откроется диалоговое окно с настройками генерации.

В диалоговом окне пользователю доступны следующие параметры:

- 1) Количество признаков - число независимых переменных в генерируемом наборе данных
  - 2) Количество наблюдений - общее число строк в генерируемом наборе данных
- После указания параметров, комплекс генерирует данные и сохраняет их в файл. Данные генерируются с использованием равномерного распределения в диапазоне от 1.0 до 100.0.

**4.1.1.1.4** После генерации данных пользователю предоставляется возможность просмотреть выборку из первых 25 объектов из сгенерированного набора данных

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

для оценки качества и структуры сгенерированных данных. Образец данных отображается в отдельном диалоговом окне.

**4.1.1.1.5** Комплекс должен предоставлять различные способы моделирования выбросов, включая генерацию случайного шума из распределений с “тяжелыми хвостами” и внесение контролируемых искажений в некоторые наблюдения. Эти настройки доступны через окно выбора метода регрессии.

**4.1.1.1.6** Комплекс предоставляет возможность выполнения статистического анализа загруженных или сгенерированных данных. При выборе кнопки для подсчета статистик выводится диалоговое окно, в котором пользователь может указать путь к CSV-файлу с данными.

После выбора файла комплекс вычисляет и отображает следующие статистические характеристики для целевой переменной:

- 1) Минимальное значение
- 2) Максимальное значение
- 3) Среднее значение
- 4) Медиана
- 5) Дисперсия

#### **4.1.1.2** Методы обработки и анализа данных

##### **4.1.1.2.1** Методы регрессионного анализа

###### **4.1.1.2.1.1** Классический метод регрессии

Пользователю предлагается воспользоваться традиционным методом регрессионного анализа – методом наименьших квадратов, который обычно чувствителен к наличию аномальных наблюдений в данных.

###### **4.1.1.2.1.2** Робастные методы регрессии

Пользователю предоставляются различные устойчивые методы регрессии, которые менее чувствительны к аномальным наблюдениям:

- 1) Методы М-оценок – используют различные функции потерь для уменьшения влияния выбросов на модель.
- 2) Методы на основе медианы – используют медианные оценки для обеспечения устойчивости к аномалиям.
- 3) Методы на основе наименьших абсолютных отклонений – минимизируют сумму абсолютных значений отклонений.

При создании модели пользователь может выбрать тип регрессии и настроить

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ее параметры, такие как:

- 1) Параметр чувствительности для робастных методов
- 2) Количество итераций для градиентных методов обучения
- 3) Скорость обучения для методов на основе градиентного спуска

#### 4.1.1.2.2 Методы обнаружения аномалий

Комплекс реализует различные алгоритмы машинного обучения для обнаружения и устранения аномальных наблюдений из набора данных.

При создании регрессионного метода пользователь может выбрать метод обнаружения аномалий, который будет применен к данным перед выполнением регрессионного анализа.

##### 4.1.1.2.2.1 Методы на основе плотности распределения

Идентифицируют аномалии путем анализа плотности распределения данных и поиска наблюдений с низкой вероятностью:

- 1) Методы оценки плотности распределения с использованием ядерных функций
- 2) Методы на основе кластеризации, выявляющие наблюдения, находящиеся далеко от основных кластеров

##### 4.1.1.2.2.2 Методы на основе расстояний

Выявляют аномалии на основе расстояний между наблюдениями

- 1) Методы поиска ближайших соседей
- 2) Методы, использующие пространственные разделения данных

##### 4.1.1.2.2.3 Метод на основе случайных разделений пространства признаков

#### 4.1.1.2.3 Визуализация и сохранение результатов

##### 4.1.1.2.3.1 Графическая визуализация результатов анализа

После выполнения анализа результаты автоматически отображаются в виде графиков, демонстрирующих зависимость ошибки от уровня шума для выбранных комбинаций регрессионных моделей и методов обнаружения аномалий.

Генерируются графики, содержащие легенды, идентифицирующие каждую серию данных на графике с указанием используемого метода обнаружения аномалий и типа распределения шума, для каждого метода регрессионной модели и метода обнаружения аномалий, позволяя наглядно сравнить их устойчивость к различным уровням и типам шума.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

По оси X отображается количество аномальных наблюдений от общего объема данных. По оси Y отображается метрика ошибки в виде среднеквадратической ошибки.

Каждая точка на графике представляет собой усредненный результат нескольких экспериментов с одинаковым уровнем шума.

#### 4.1.1.2.3.2 Расчет и сохранение метрик качества регрессии

Комплекс вычисляет набор различных метрик для комплексной оценки качества регрессионных моделей:

- 1) Средняя абсолютная ошибка (MAE) - среднее абсолютных значений разностей между прогнозируемыми и фактическими значениями
- 2) Среднеквадратическая ошибка (MSE) - среднее квадратов разностей между прогнозируемыми и фактическими значениями
- 3) Корень из среднеквадратической ошибки (RMSE) - квадратный корень из MSE
- 4) Средняя абсолютная процентная ошибка (MAPE)
- 5) Симметричная средняя абсолютная процентная ошибка (SMAPE)

Вычисленные метрики сохраняются в структурированном формате в выходной JSON файл. Пример организации подобного файла приведен в Приложении 2.

#### 4.1.1.2.3.3 Расчет и сохранение метрик качества обнаружения аномальных наблюдений

Комплекс позволяет оценивать эффективность различных методов обнаружения аномалий путем расчета стандартных метрик качества классификации. Эти метрики помогают количественно оценить способность алгоритма правильно идентифицировать аномальные наблюдения.

Вычисляемые метрики качества обнаружения аномалий:

- 1) Точность (Precision) - доля правильно идентифицированных аномалий среди всех наблюдений, отмеченных как аномальные
- 2) Полнота (Recall) - доля правильно идентифицированных аномалий среди всех фактических аномальных наблюдений
- 3) F1-мера - гармоническое среднее между точностью и полнотой

Метрики качества обнаружения аномальных наблюдений сохраняются в JSON-файл вместе с метриками качества регрессионных моделей. Для каждого метода обнаружения аномалий и каждого уровня шума сохраняются

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

все рассчитанные метрики. Пример организации подобного файла приведён в Приложении 2.

#### 4.1.2. Требования к организации входных данных

Входные данные должны быть представлены в формате CSV с использованием точки с запятой в качестве разделителя. Первой строкой файла должны быть названия столбцов (заголовки). Каждое имя столбца должно быть уникальным и не содержать специальных символов. Данные внутри столбцов должны быть однородными: для числовых переменных допустимы только значения типа float или int. Пропуски данных недопустимы, при обнаружении пропусков пользователю будет предложено удалить признаки с пропусками. Данные не должны содержать дублирующихся строк. При обнаружении дублирования пользователю будет предложено удалить их. Размер файла ограничен 1 Гб. При превышении лимита пользователь получает предупреждение с предложением уменьшить объем данных.

#### 4.1.3. Требования к организации выходных данных

Все результаты работы комплекса должны сохраняться в CSV, JSON или PNG форматах. Результаты сравнения методов сохраняются в JSON-формате, как описано в пунктах 4.1.1.2.3.3. и 4.1.1.2.3.2. Графики, построенные в процессе анализа, будут сохранены в формате PNG.

#### 4.1.4. Требования к интерфейсу

##### 4.1.4.1. Общие

требования

Интерфейс программного комплекса должен быть интуитивно понятным и удобным для пользователя, обеспечивая доступ ко всем функциям через понятные и логически организованные элементы управления. Графический интерфейс (GUI) должен поддерживать управление с использованием мыши и клавиатуры.

##### 4.1.4.2. Требования к интерфейсу работы с данными

При запуске комплекса на главном экране должны отображаться кнопки “Create model”, при нажатии которой должно открываться диалоговое окно, позволяющее выбрать тип регрессионной модели и настроить ее параметры, “Dump models” для сохранения конфигураций созданных моделей, “Run on models”, при нажатии которой должен запускаться анализ на основе созданных моделей, “Generate data” для генерации синтетических данных и “Calculate

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

statistics” для расчета статистик целевой переменной.

#### 4.1.4.3. Требования к интерфейсу визуализации результатов

После запуска анализа результаты должны автоматически визуализироваться в виде графиков, показывающих зависимость ошибки от уровня шума для каждой комбинации регрессионной модели и метода обнаружения аномалий. Графики должны отображаться в отдельном окне и включать заголовок с указанием типа регрессионной модели и метода обнаружения аномалий, подписи осей, точки, представляющие усредненные результаты экспериментов для каждого уровня шума и легенду, идентифицирующую каждую серию данных на графике с указанием используемого метода обнаружения аномалий и типа распределения шума.

#### 4.1.4.4. Требования к интерфейсу анализа и сохранения результатов

Интерфейс должен обеспечивать возможность просмотра графиков зависимости ошибки от уровня шума, метрик качества регрессионных моделей и метрик качества обнаружения аномалий. Все результаты анализа должны автоматически сохраняться в структурированном формате JSON, содержащем информацию о модели регрессии, информацию о методе обнаружения аномалий, метрики качества регрессии для каждого уровня шума, метрики качества обнаружения аномалий для каждого уровня шума. Файлы результатов должны сохраняться с именами, содержащими информацию о типе регрессионной модели и методе обнаружения аномалий, для удобства последующего анализа.

## 4.2. Требования к надежности

При создании модели, если указанный путь к файлу данных не существует, программа должна выводить информативное сообщение об ошибке и оставлять поле пути пустым. Если файл данных имеет неправильный формат, программа должна выводить соответствующее сообщение об ошибке. При невозможности сохранения результатов, программа должна информировать пользователя о проблеме.

При вводе недопустимых значений параметров, интерфейс должен блокировать ввод

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

**RU.17701729.11.04-01 ТЗ 01-1**

таких значений или выводить предупреждение. При попытке запуска анализа без созданных моделей, программа должна информировать пользователя о необходимости создания моделей. При возникновении ошибок в процессе вычислений, программа должна обрабатывать эти ошибки и продолжать работу с оставшимися моделями

При сбоях программного комплекса состояние программы и результаты анализа должны быть сохранены в памяти для повторной попытки.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

### 4.3. Условия эксплуатации

Требования к условиям эксплуатации программного продукта совпадают с требованиями эксплуатации устройства, используемого для работы с программным обеспечением. Программа предназначена для работы на персональных компьютерах и серверах, соответствующих минимальным системным требованиям, указанным в документации. Специальных требований к условиям эксплуатации приложения, таких как температура, влажность или другие физические параметры, не предъявляется.

### 4.4. Требования к составу и параметрам технических средств

#### 4.4.1. Требования к клиентскому оборудованию

Рекомендуемые требования к клиентскому оборудованию для корректной работы приложения:

1. Процессор с 4 или более ядрами (AMD Ryzen 5 / Intel Core i5 или другие);
2. 8 ГБ оперативной памяти или больше;
3. 2 ГБ видеопамяти или больше;
4. Монитор разрешением не менее 1920x1080.
5. Наличие SSD-диска с не менее 8 ГБ свободного пространства.

Минимальные требования к клиентскому оборудованию для работы приложения:

1. Процессор Intel Core i3-5100f;
2. 4 ГБ оперативной памяти;
3. 2 ГБ видеопамяти;
4. Монитор разрешением 1280x720.
5. Жесткий диск с не менее 5 ГБ свободного пространства.

Общие требования к клиентскому оборудованию для работы приложения:

1. Мышь или совместное указывающее устройство;
2. Клавиатура;
3. ОС: Windows (10 и выше), Linux (Ubuntu 20.04 и выше) или macOS версии не ниже 11 “Big Sur”;

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



#### 4.5. Требования к информационной и программной совместимости

Приложение должно быть написано в соответствии со стандартом языка C++17. Для сборки проекта используется система сборки CMake версии не ниже 3.14. Должен использоваться компилятор GCC версии не ниже 10.5 или Clang версии не ниже 13.0.0. Для управления версиями приложения используется система контроля версий Git. Версия языка Python должна быть не ниже 3.9.

#### 4.6. Требования к маркировке и упаковке

Программа распространяется в виде электронного пакета, содержащего программную документацию, приложение (исполняемые файлы и прочие необходимые для работы файлы).

#### 4.7. Требования к транспортировке

Особых требований к транспортировке и хранению не предъявляются.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

### 5.1. Предварительный состав программной документации

1. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Техническое задание (ГОСТ 19.201-78<sup>[7]</sup>).
2. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Пояснительная записка (ГОСТ 19.404-79<sup>[11]</sup>).
3. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Программа и методика испытаний (ГОСТ 19.301-79<sup>[12]</sup>).
4. «Разработка программного комплекса для исследования влияния аномальных наблюдений на —точность прогнозирования в регрессионных моделях». Текст программы (ГОСТ 19.401-78<sup>[13]</sup>).
5. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Руководство оператора (ГОСТ 19.505-79<sup>[14]</sup>).
6. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Руководство программиста (ГОСТ 19.504-79<sup>[15]</sup>).

### 5.2. Специальные требования к программной документации

Программная документация должна быть выполнена в соответствии с ГОСТ 19.106-78<sup>[6]</sup> и ГОСТами к каждому виду документа.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

### 6.1 Ориентировочная экономическая эффективность

В рамках проекта расчет экономической эффективности программного продукта не производился.

### 6.2 Предполагаемая потребность

Регрессионные модели машинного обучения широко используются в разных областях бизнеса и науки. Данные, на которых эти модели обучаются не всегда свободны от выбросов, что, в свою очередь, создает опасность ухудшения качества предсказаний и снижения эффективности использования регрессионных моделей. Разрабатываемый программный комплекс может использоваться как для оценки просадок качества при обучении на богатых выбросами выборках, так и средства борьбы с выбросами (если нельзя выбросить и/или заменить аномальные наблюдения из выборки, мы должны обеспечить возможность обеспечить устойчивость модели к этим аномальным наблюдениям).

### 6.3 Экономические преимущества разработки по сравнению с отечественными и зарубежными образцами или аналогами

Первое достоинство комплекса заключается в предоставлении средств для анализа влияния выбросов на качество моделей. Пользователь сможет легко проверить, как именно изменится качество его модели при включении части выбросов или их всех. Таким образом, пользователь получает механизм гибкой настройки модели и данных, позволяющий добиться именно того результата, к которому он, пользователь, идет. Второе достоинство заключается в предоставлении комплексом средств, позволяющих бороться с выбросами. Присутствуют как средства, разработанные на основе методов математической статистики и качество которых имеет доказательство, так и методы машинного обучения, направленные на обнаружение аномальных наблюдений.

Помимо пользы в работе с применением технологией машинного обучения, можно выделить и образовательную пользу. Программный комплекс может использоваться в качестве обучающего материал для демонстрации влияния выбросов на качество регрессионных моделей и средств создания устойчивых к выбросам моделей.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 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 учётом ГОСТ 19.102-77<sup>[2]</sup>.

Стадия разработки	Этап работ	Содержание работ	Сроки выполнения
Техническое задание	Обоснование необходимости разработки программы	<p>Постановка задачи</p> <p>Сбор исходных материалов.</p> <p>Выбор и обоснование критериев эффективности и качества разрабатываемой программы.</p>	04.11.2024
	Научно-исследовательские работы	<p>Обоснование возможности решения поставленной задачи.</p> <p>Предварительный выбор методов решения задач.</p> <p>Определение требований к техническим и программным средствам</p>	21.11.2024

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

	Разработка и утверждение технического задания	<p>Определение требований к программе.</p> <p>Определение стадий, этапов и сроков разработки программы и документации на неё.</p> <p>Согласование и утверждение технического задания.</p> <p>Загрузка согласованного технического задания в SmartLMS</p>	04.12.2024
Рабочий проект	Разработка программы	<p>Предварительная разработка структуры программы</p> <p>Разработка учебных материалов</p>	27.02.2025
	Разработка программной документации	Разработка программных документов в соответствии с требованиями ГОСТ 19.101-77 <sup>[1]</sup> .	03.03.2025

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Внедрение	Подготовка и передача программы	Подготовка программы и программной документации для презентации и защиты  Представление разработанного программного продукта научному руководителю и получение отзыва  Представление разработанного программного продукта научному руководителю и получение отзыва.  Загрузка Пояснительной записки в систему Антиплагиат через ЛМС НИУ ВШЭ.  Загрузка материалов курсового проекта в ЛМС  Защита программного продукта комиссии.	15.03.2025
-----------	---------------------------------	---	------------

## 7.2. Сроки разработки и исполнители

Разработка программного продукта должна быть завершена к XX.XX.2025. (дата защиты курсовой работы). Исполнители – Знатнов Егор и Панкратов Степан, студенты группы БПИ221 факультета компьютерных наук НИУ ВШЭ

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

**8. ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ**

Контроль и приемка разработки осуществляются в соответствии с документом «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Программа и методика испытаний» и пунктом 5.2 технического задания.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## ПРИЛОЖЕНИЕ 1

**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 19.602-78: Правила дублирования, учета и хранения программных документов, выполненных печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.603-78: Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 19.604-78: Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
11. ГОСТ 19.404-79: Пояснительная записка. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. ГОСТ 19.301-79: Программа и методика испытаний. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



13. ГОСТ 19.401-78: Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
14. ГОСТ 19.505-79: Руководство оператора. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
15. ГОСТ 19.504-79: Руководство программиста. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
16. Maronna R. A. et al. Robust statistics: theory and methods (with R). – John Wiley & Sons, 2019. [Электронный ресурс] /Режим доступа: приватный (дата обращения: 26.11.2024).
17. Chang W. H. et al. High-breakdown rank regression //Journal of the American Statistical Association. – 1999. – Т. 94. – №. 445. – С. 205-219. [Электронный ресурс] /Режим доступа: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473836> Свободный (дата обращения: 26.11.2024)
18. Hettmansperger T. P., McKean J. W. Robust nonparametric statistical methods. – CRC press, 2010. [Электронный ресурс] /Режим доступа: приватный (дата обращения: 26.11.2024)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## ПРИЛОЖЕНИЕ 2

```

{
  "results": [
    {
      "model": "LSM",
      "mlmodel": "None",
      "noise_type": "Normal",
      "noise_params": {
        "param1": 0.0,
        "param2": 1.0
      },
      "data_path": "sample.csv",
      "num_features": 3,
      "metrics_by_noise_level": [
        {
          "noise_level": 0,
          "metrics": {
            ...
          },
          "detection_metrics": {
            "f1_score": 0.0,
            "precision": 0.0,
            "recall": 0.0,
          },
          "num_experiments": 10
        },
        //...
      ]
    },
    //...
  ]
}

```

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

### ТЕРМИНЫ И СОКРАЩЕНИЯ

1. Модель – объект, представляющий собой набор весов, показывающих степень влияния некоторого признака в наборе данных. Модель используется для обнаружения зависимостей в наборе данных.
2. Метрика – некоторая мера, позволяющая измерить качество некоторой модели путем сравнения результатов применения модели с данными, описанными в наборе данных.
3. Регрессия – задача математической статистики, суть которой заключается в предсказании некоторой числовой величины по известным данным-признакам.
4. Признак – поле в наборе данных, которое используется как для предсказания величины-цели.
5. Целевая переменная (цель) – поле в наборе данных, значение которого модель стремится предсказать на основе признаков.
6. Аномальное наблюдение – объект в наборе данных, признаки которого многократно превосходят средние показатели других объектов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.10-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Лист регистрации изменений
----------------------------

[illegible]