

**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**

Программа и методика испытаний

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.11.04-01 ПМИ 03-1

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

УТВЕРЖДЕН

RU.17701729.11.04-01 ПМИ 03-1

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

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

**Программа и методика испытаний
RU.17701729.11.04-01 ПМИ 03-1**

Листов 31

АННОТАЦИЯ

Программа и методика испытаний — это документ, в котором содержится информация о программном продукте, а также полное описание приемочных испытаний для данного программного продукта.

Настоящая Программа и методика испытаний для «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях» содержит следующие разделы:

«Объект испытаний», «Цель испытаний», «Требования к программе», «Требования к программной документации», «Средства и порядок испытаний», «Методы испытаний», «Приложения».

В разделе «Объект испытаний» указано наименование, краткая характеристика и назначение программы.

В разделе «Цель испытаний» указана цель проведения испытаний.

Раздел «Требования к программе» содержит основные требования к программе, которые подлежат проверке во время испытаний (требования к функционалу).

Раздел «Требования к программным документам» содержит состав программной документации, которая представляется на испытания.

Раздел «Средства и порядок испытаний» содержит информацию о технических и программных средствах, которые следует использовать во время испытаний, а также порядок этих испытаний.

Раздел «Методы испытаний» содержит информацию об используемых методах испытаний.

Настоящий документ разработан в соответствии с требованиями:

- 1) ГОСТ 19.101-77 Виды программ и программных документов [1];
- 2) ГОСТ 19.102-77 Стадии разработки [2];
- 3) ГОСТ 19.103-77 Обозначения программ и программных документов [3]; 4) ГОСТ 19.104-78 Основные надписи [4];
- 5) ГОСТ 19.105-78 Общие требования к программным документам [5];
- 6) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];
- 7) ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению [10].

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Изменения к данному документу оформляются согласно ГОСТ 19.603-78 [8], ГОСТ 19.604-78 [9].

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

ГЛОССАРИЙ.....	7
1. ОБЪЕКТ ИСПЫТАНИЙ	8
1.1. Наименование программы.....	8
1.2. Краткая характеристика области применения	8
2. ЦЕЛЬ ИСПЫТАНИЙ.....	9
2.1. Документы, на основании которых ведется разработка	9
3. ТРЕБОВАНИЯ К ПРОГРАММЕ	10
3.1. Требования к функциональным характеристикам	10
3.1.1. Состав выполняемых функций.....	10
3.1.1.1. Создание и управление регрессионными моделями	10
3.1.1.2. Генерация и обработка данных с аномалиями.....	11
3.1.1.3. Методы очистки данных от аномалий	13
3.1.1.4. Оценка влияния аномалий на качество регрессионных моделей	14
3.1.2. Организация хранения данных.....	16
3.1.3. Организация выходных данных	17
3.1.4. Требования к интерфейсу.....	18
3.2. Требования к надежности	20
4. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ.....	22
4.1. Состав программной документации	22
4.2. Специальные требования к программной документации.....	22
5. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ	24
5.1. Технические средства, используемые во время испытаний	24
5.2. Программные средства, используемые во время испытаний	24
5.3. Порядок проведения испытаний	25
6. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ	26
6.1. Проверка требований к технической документации.....	26
6.2. Проверка требований к функциональным характеристикам.....	26
6.2.1. Модуль пользовательского интерфейса	26
6.2.1.1. Проверка создания моделей	26

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6.2.1.2.	Проверка сохранения моделей.....	26
6.2.1.3.	Проверка генерации данных.....	26
6.2.1.4.	Проверка запуска анализа.....	27
6.2.2.	Модуль статистических методов.....	27
6.2.2.1.	Метод наименьших квадратов (LSM)	27
6.2.2.2.	Метод наименьших абсолютных отклонений (LAD):	27
6.2.2.3.	Робастные методы регрессии (Huber, Tukey).....	27
6.2.2.4.	Проверка реализации метода Тейла-Сена (TheilSen).....	27
6.2.3.	Модуль машинного обучения	27
6.2.4.	Модель визуализации.....	28
6.2.4.1.	Проверка генерации графиков	28
6.2.4.2.	Проверка сохранения результатов	28
6.3.	Проверка требований к надёжности	28
6.3.1.	Проверка устойчивости программы к некорректным входным данным	28
6.3.2.	Проверка параллельной обработки данных.....	28
6.3.3.	Проверка работы с памятью.....	28
СПИСОК ИСТОЧНИКОВ		29
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ		32

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

ГЛОССАРИЙ

1) **Регрессионная модель** – математический метод прогнозирования, устанавливающий зависимость между целевой переменной и одним или несколькими признаками.

2) **Аномальные наблюдения** – точки данных, которые значительно отклоняются от остальных наблюдений в наборе данных и могут негативно влиять на точность прогнозирования.

3) **JSON** – легкий формат обмена данными, используемый для хранения конфигураций моделей и параметров экспериментов.

4) **CSV** – формат хранения табличных данных, где значения разделены определенным символом.

5) **MVC (Model-View-Controller)** – архитектурный паттерн, используемый для разделения логики приложения и представления данных.

6) **Асинхронные вычисления** – метод параллельного выполнения задач для повышения производительности, особенно при проведении множества экспериментов.

7) **Целевая переменная (цель)** – поле в наборе данных, значение которого модель стремится предсказать на основе признаков.

8) **Признак** – поле в наборе данных, которое используется как для предсказания величины-цели.

9) **Метрики качества** – показатели, используемые для оценки точности регрессионных моделей.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

1. ОБЪЕКТ ИСПЫТАНИЙ

1.1. Наименование программы

Наименование программы – «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях».

Наименование программы на английском языке – «Development of a Software Package to Study the Influence of Outliers on the Prediction Accuracy in Regression Models».

Наименование программы для пользователя – «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.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

2. ЦЕЛЬ ИСПЫТАНИЙ

2.1. Документы, на основании которых ведется разработка

Целью испытаний является проверка корректности выполнения программой функций, изложенных в п. 4 «Требования к программе» настоящего Технического задания из комплекта документации в соответствии с ЕСПД (Единой системой программной документации).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

3. ТРЕБОВАНИЯ К ПРОГРАММЕ

3.1. Требования к функциональным характеристикам

3.1.1. Состав выполняемых функций

3.1.1.1. Создание и управление регрессионными моделями

Цель:

Разработка и реализация различных типов регрессионных моделей с возможностью настройки их параметров для исследования их устойчивости к аномальным наблюдениям.

Задачи:

- 1) Реализация классического метода наименьших квадратов (LSM) для получения базовой модели регрессии.
- 2) Реализация робастных методов регрессии (Huber, Tukey) с настраиваемыми параметрами чувствительности к выбросам.
- 3) Реализация метода наименьших абсолютных отклонений (LAD) для минимизации суммы абсолютных разностей между наблюдаемыми и предсказанными значениями.
- 4) Реализация метода Тейла-Сена (TheilSen) для оценки медианных наклонов между всеми парами точек данных.
- 5) Обеспечение возможности настройки параметров моделей через графический интерфейс пользователя.

Требования к моделям:

Модели должны обеспечивать корректные результаты на чистых данных без аномалий.

Параметры моделей должны быть легко изменяемы.

Все модели должны возвращать результаты в едином формате для корректного сравнения.

Методы реализации:

Использование библиотеки Eigen для эффективных матричных вычислений.

Применение методов оптимизации на основе градиентного спуска для обучения моделей.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Абстрагирование общих функций в базовые классы для обеспечения единообразия интерфейса моделей.

Процедура:

Создание базового класса StatsMethod с общими методами для всех регрессионных моделей.

Реализация специфических методов для каждого типа регрессии.

Интеграция моделей с графическим интерфейсом для удобной настройки параметров.

Разработка механизма сохранения параметров модели в JSON-формате для последующего использования.

Контрольные точки:

Корректная работа каждой регрессионной модели на входных данных.

Подтверждение возможности настройки всех параметров через пользовательский интерфейс.

Успешное сохранение и загрузка параметров моделей в/из JSON-файла.

Критерии приемки:

Все реализованные модели корректно обучаются на данных и дают ожидаемые результаты.

Пользовательский интерфейс обеспечивает интуитивно понятный способ настройки параметров моделей.

Система успешно сохраняет и загружает конфигурации моделей в JSON-формате.

3.1.1.2. Генерация и обработка данных с аномалиями

Цель:

Создание механизмов для генерации синтетических данных с контролируемым уровнем и типом аномалий для тестирования устойчивости регрессионных моделей.

Задачи:

- 1) Разработка генератора синтетических данных с заданным количеством признаков и наблюдений.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

- 2) Реализация механизмов добавления шума различных типов к данным.
- 3) Обеспечение возможности контроля процента аномальных наблюдений зашумления данных.
- 4) Реализация функциональности для загрузки внешних датасетов в формате CSV.

Требования к генерации данных:

Возможность создания наборов данных с различным количеством признаков и наблюдений.

Четкий контроль над процентом и типом добавляемых аномалий.

Методы реализации:

Применение статистических распределений для генерации шума различных типов.

Разработка простого интерфейса для настройки параметров генерации данных.

Процедура:

Создание основных функций для генерации чистых синтетических данных.

Реализация классов для различных типов распределений шума.

Разработка механизма применения шума к определенному проценту наблюдений.

Интеграция с пользовательским интерфейсом для удобной настройки параметров генерации.

Контрольные точки:

Корректная генерация синтетических данных с заданными параметрами.

Успешное добавление шума различных типов к данным.

Подтверждение контроля над процентом аномальных наблюдений.

Критерии приемки:

Генератор создает синтетические данные с заданным количеством признаков и наблюдений.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Система успешно добавляет шум различных типов с контролем над уровнем аномальности.

Пользовательский интерфейс обеспечивает простой способ настройки параметров генерации.

3.1.1.3. Методы очистки данных от аномалий

Цель:

Разработка и реализация алгоритмов машинного обучения для обнаружения и удаления аномальных наблюдений из наборов данных для повышения точности регрессионных моделей.

Задачи:

- 1) Реализация метода Isolation Forest для обнаружения аномалий на основе изоляции наблюдений.
- 2) Реализация метода DBSCAN для кластеризации и выявления аномалий как точек с низкой плотностью.
- 3) Реализация метода Kernel Density Estimation для определения вероятностной плотности распределения и выявления аномалий.
- 4) Реализация метода k-ближайших соседей для обнаружения локальных аномалий.
- 5) Обеспечение возможности применения выбранного метода для предобработки данных перед обучением регрессионных моделей.

Требования к методам очистки:

Методы должны успешно выявлять аномалии различных типов.

Основные параметры методов должны быть настраиваемыми или заранее установленными.

Методы должны быть легко интегрируемы в общий процесс анализа данных.

Методы реализации:

Разработка классов для каждого алгоритма обнаружения аномалий с единым интерфейсом.

Применение оптимизированных структур данных для повышения производительности.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Использование многопоточных вычислений для ускорения процесса обнаружения аномалий.

Процедура:

Разработка базового класса для методов обнаружения аномалий.

Реализация специфических алгоритмов для каждого метода.

Создание класса DataDeNoiser для объединения функциональности и предоставления единого интерфейса.

Интеграция с пользовательским интерфейсом и регрессионными моделями.

Контрольные точки:

Успешное обнаружение искусственно добавленных аномалий с помощью каждого метода.

Подтверждение повышения точности регрессионных моделей после удаления аномалий.

Сравнительный анализ эффективности различных методов очистки данных.

Критерии приемки:

Методы успешно обнаруживают и удаляют аномалии различных типов.

Очистка данных приводит к улучшению точности регрессионных моделей.

Пользовательский интерфейс обеспечивает простой способ выбора и настройки методов очистки.

3.1.1.4. Метрики качества регрессии и обнаружения аномалий

Цель:

Разработка методологии и инструментов для систематического исследования и количественной оценки влияния аномальных наблюдений на точность и надежность различных регрессионных моделей.

Задачи:

1) Реализация механизма проведения экспериментов с различными уровнями шума для каждой регрессионной модели.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

- 2) Разработка метрик для оценки качества методов обнаружения аномальных наблюдений и регрессии на зашумленных данных.
- 3) Создание системы визуализации результатов экспериментов в виде графиков зависимости ошибки от уровня шума.
- 4) Обеспечение возможности сравнения эффективности различных методов очистки данных при разных типах и уровнях аномалий.

Требования к оценке:

Эксперименты должны проводиться по единой методологии для обеспечения сравнимости результатов.

Результаты должны предоставлять четкую картину влияния аномалий на каждую модель.

Визуализации должны ясно демонстрировать зависимости и тренды.

Методы реализации:

Автоматизированное проведение серий экспериментов с различными параметрами.

Применение асинхронных вычислений для параллельного выполнения экспериментов.

Использование Python с библиотекой matplotlib для создания наглядных визуализаций.

Процедура:

Разработка класса для автоматизированного проведения экспериментов с заданными параметрами.

Реализация вычисления метрик качества для оценки моделей.

Создание системы сохранения результатов экспериментов.

Разработка механизма генерации графиков для визуализации результатов.

Контрольные точки:

Успешное проведение серии экспериментов с различными уровнями шума для всех моделей.

Корректное вычисление метрик качества и сохранение результатов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Создание информативных визуализаций зависимости ошибки от уровня шума.

Критерии приемки:

Система успешно проводит эксперименты с разными типами и уровнями аномалий.

Метрики качества корректно отражают влияние аномалий на точность моделей.

3.1.2. Организация хранения данных

Цель:

Структурировать и систематизировать хранение данных для регрессионного анализа и исследования влияния аномалий на точность моделей.

Задачи:

- 1) Организовать эффективное хранение данных для экспериментов в формате CSV.
- 2) Обеспечить структуру данных, оптимальную для обработки различными регрессионными методами.
- 3) Разработать формат хранения конфигурации моделей и параметров экспериментов в JSON.

Требования к характеристикам данных:

Данные должны иметь единый формат разделителей и структуру для корректной обработки.

Файлы данных должны быть легко читаемы и модифицируемы.

Наборы данных должны содержать все необходимые признаки и целевые переменные.

Методы организации хранения:

Хранение данных регрессии в формате CSV с разделителями ";" для обеспечения совместимости и удобства анализа.

Организация структуры файла с набором признаков в каждой строке и целевым значением в последнем столбце.

Сохранение конфигураций моделей в формате JSON с четкой структурой, включающей тип модели, параметры и настройки шума.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Контрольные точки:

Создание структурированных CSV-файлов с данными для проведения экспериментов.

Формирование корректных JSON-конфигураций для задания параметров моделей и экспериментов.

Проверка доступности и целостности данных для различных методов регрессии и алгоритмов обнаружения аномалий.

Критерии приемки:

Эффективность и удобство системы хранения данных для проведения исследований.

Корректная структура файлов данных и конфигураций, обеспечивающая беспрепятственное использование в анализе.

3.1.3. Организация выходных данных**Цель:**

Обеспечить представление результатов исследования влияния аномалий на точность регрессионных моделей в наглядном и информативном формате.

Задачи:

- 1) Генерация графических представлений зависимости ошибки от уровня шума для различных моделей.
- 2) Сохранение результатов экспериментов в структурированном формате для последующего анализа.
- 3) Обеспечение возможности визуального сравнения эффективности различных методов устойчивой регрессии.

Требования к выходным данным:

Данные должны наглядно демонстрировать влияние аномалий на точность прогнозирования.

Результаты должны быть представлены в формате, удобном для сравнения различных моделей и методов очистки.

Результаты анализа должны быть легко интерпретируемы даже для пользователей без специальных знаний в статистике.

Методы реализации:

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Генерация графиков для каждой выбранной комбинации регрессионной модели и метода очистки данных.

Сохранение визуализаций в формате PNG с четкими заголовками и обозначением осей.

Формирование структурированных конфигурационных файлов для графиков в формате JSON.

Процедура:

Проведение экспериментов с различными уровнями шума для каждой регрессионной модели.

Формирование конфигурации графика с указанием названия, осей и параметров отображения.

Сохранение результатов в виде графических изображений с говорящими именами, включающими тип модели и метод очистки.

Контрольные точки:

Успешное создание и сохранение графиков для всех комбинаций моделей и методов.

Информативность и читаемость полученных визуализаций.

Корректная маркировка и обозначение элементов на графиках.

Критерии приемки:

Графики корректно отображают зависимость ошибки от уровня шума для разных моделей.

Визуализации имеют четкие заголовки, оси и детали, облегчающие интерпретацию.

Результаты сохраняются в форматах, удобных для последующего анализа и включения в отчеты.

3.1.4. Требования к интерфейсу

Цель:

Разработать интуитивно понятный графический интерфейс для управления процессом создания моделей, генерации данных и проведения исследований влияния аномалий.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Задачи:

Разработка интерфейса для создания и настройки регрессионных моделей различных типов.

Реализация функциональности для генерации данных с заданными параметрами.

Обеспечение возможности запуска экспериментов и визуализации результатов.

Требования к интерфейсу:

Интерфейс должен быть понятным и простым в использовании.

Интерфейс должен предоставлять доступ ко всем функциям программы.

Пользователь должен получать обратную связь о результатах выполненных операций.

Методы реализации:

Использование библиотеки Qt для создания современного и кроссплатформенного графического интерфейса.

Применение модели MVC для четкого разделения логики и представления.

Реализация диалоговых окон для ввода параметров моделей и экспериментов.

Процедура:

Разработка основного окна приложения с кнопками для основных функций: создание модели, генерация данных, запуск анализа.

Реализация диалоговых окон для ввода параметров моделей и данных.

Интеграция с вычислительными модулями для проведения экспериментов.

Создание механизма отображения результатов анализа в виде графиков.

Контрольные точки:

Функциональность создания и настройки моделей через интерфейс.

Корректная работа генерации данных с пользовательскими параметрами.

Возможность сохранения моделей и запуска экспериментов.

Визуализация результатов анализа.

Критерии приемки:

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Интерфейс позволяет создавать и настраивать все типы поддерживаемых моделей.

Генерация данных работает корректно с разными параметрами.

Эксперименты запускаются и выполняются без ошибок.

Результаты анализа отображаются в понятном и информативном виде.

3.2. Требования к надежности

Цель:

Обеспечить стабильную и надежную работу программного комплекса при проведении исследований влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях.

Требования к устойчивости:

Программа должна корректно обрабатывать некорректные входные данные, предоставляя пользователю информативные сообщения об ошибках.

При отсутствии файлов данных программа должна выдавать соответствующее предупреждение и предлагать альтернативные действия.

Система должна сохранять работоспособность при неправильном формате входных файлов, обрабатывая исключения без аварийного завершения.

Требования к обработке ошибок:

Все потенциальные ошибки должны быть перехвачены и обработаны с предоставлением пользователю понятной обратной связи.

При невозможности выполнения операции система должна предлагать альтернативные варианты действий или четкие инструкции по устранению проблемы.

Ошибки в процессе вычислений должны логироваться для последующего анализа.

Требования к производительности:

Программа должна эффективно использовать многопоточность для ускорения процесса вычислений.

Система должна минимизировать использование памяти при работе с большими наборами данных.

Требования к восстановлению:

При аварийном завершении программа должна сохранять промежуточные

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

результаты и обеспечивать возможность восстановления работы.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

4. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

4.1. Состав программной документации

4.1.1. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Техническое задание (ГОСТ 19.201-78) [7];

4.1.2. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Программа и методика испытаний (ГОСТ 19.301-78) [10];

4.1.3. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Пояснительная записка (ГОСТ 19.404-79) [11];

4.1.4. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Руководство программиста (ГОСТ 19.504-79) [12];

4.1.5. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Текст программы (ГОСТ 19.401-78) [13];

4.1.6. «Разработка программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях». Руководство оператора (ГОСТ 19.505-79) [14];

4.2. Специальные требования к программной документации

1) Документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 и ГОСТами к каждому виду документа (см. п. 5.1).

2) Пояснительная записка должна быть загружена в систему Антиплагиат через LMS «НИУ ВШЭ». Лист, подтверждающий загрузку пояснительной записки, сдается в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы.

3) Вся документация также воспроизводится в печатном виде, она должна быть подписана академическим руководителем образовательной программы 09.03.04 «Программная инженерия», руководителем разработки и исполнителем перед

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

сдачей курсовой работы в учебный офис, не позже одного дня до защиты.

4) Документация также сдается в электронном виде в формате .pdf или .docx, а программа – в системе контроля версий Git.

5) Все документы перед защитой курсовой работы должны быть загружены в информационно-образовательную среду НИУ ВШЭ LMS (Learning Management System) в личном кабинете, дисциплина – «Курсовой проект, 3 курс ПИ», одним архивом.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

5. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ

5.1. Технические средства, используемые во время испытаний

В рамках тестирования программного комплекса для исследования влияния аномальных наблюдений на точность прогнозирования в регрессионных моделях использовался следующий конфигурационный набор оборудования:

1) Операционная система: macOS Monterey, обеспечивающая высокую производительность и стабильность работы для приложений с графическим интерфейсом и вычислительных процессов.

2) Процессор: Apple M1 Pro (Macbook Pro 16 Retina), предоставляющий высокую вычислительную мощность и энергоэффективность для машинного обучения и статистического анализа.

3) Оперативная память: 16 Гб унифицированной памяти, что позволяет эффективно обрабатывать большие объемы данных и выполнять параллельные вычисления без существенных задержек.

4) Накопитель: 512 Гб SSD, обеспечивающий быстрый доступ к данным и ускоренную загрузку компонентов программы.

5) Графический процессор: Встроенный в M1 Pro GPU с 16 ядрами, способствующий эффективной визуализации результатов экспериментов.

5.2. Программные средства, используемые во время испытаний

Во время испытаний использовались следующие программные средства:

1) Среда разработки: QtCreator версии 6.0.2 для разработки и отладки приложения с графическим интерфейсом

2) Компилятор: Clang 13.0.0 для компиляции исходного кода C++

3) Библиотека Qt 5.15 для создания графического пользовательского интерфейса

4) Библиотека Eigen 3.4.0 для матричных вычислений и операций линейной алгебры

5) Библиотека nlohmann::json 3.11.3 для обработки JSON-данных

6) Python 3.9.7 для генерации графиков и визуализации результатов исследований

7) Библиотеки Python: matplotlib 3.10.3 и numpy 1.25.0 для построения графиков

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

5.3. Порядок проведения испытаний

Испытания должны проводиться в следующем порядке:

- 1) Установить и настроить все необходимые программные средства, указанные в разделе "Программные средства".
- 2) Скомпилировать исходный код приложения из репозитория с использованием QtCreator или командной строки.
- 3) Запустить скомпилированное приложение.
- 4) Провести следующие операции в пользовательском интерфейсе:
 - 4.1) Сгенерировать данные с помощью кнопки "Generate data"
 - 4.2) Создать модели для анализа с помощью кнопки "Create model"
 - 4.3) Сохранить созданные модели с помощью кнопки "Dump models"
 - 4.4) Запустить процесс анализа с помощью кнопки "Run on models"
- 5) Проверить корректность вывода результатов визуализации и анализа.
- 6) Проверить корректность вывода подсчета статистических характеристик при использовании кнопки "Calculate statistics".
- 7) Провести испытания, описанные в разделе "Методика испытаний".
- 8) Оценить точность прогнозирования различных регрессионных моделей при наличии аномальных наблюдений.
- 9) Завершить работу приложения.

Все испытания должны выполняться в соответствии с требованиями и последовательностью, описанными в методике испытаний. Результаты испытаний должны быть задокументированы для последующего анализа и оценки эффективности разработанного программного комплекса.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ

6.1. Проверка требований к технической документации

Состав программной документации проверяется наличием полного комплекта документов программной документации в системе SmartLMS и наличием всех требуемых подписей. Также проверяется соответствие документации требованиям ГОСТ.

Комплект документов полный. Все документы удовлетворяют представленным требованиям.

6.2. Проверка требований к функциональным характеристикам

6.2.1. Модуль пользовательского интерфейса

6.2.1.1. Проверка создания моделей

- 1) Проверить, что при нажатии кнопки "Create model" открывается диалоговое окно с корректными полями.
- 2) Убедиться, что все поля формы (тип модели, дельта, количество эпох, скорость обучения, тип шума, параметры шума, ML-модель) корректно обрабатываются.
- 3) Проверить, что созданная модель добавляется в таблицу моделей с правильными параметрами.

6.2.1.2. Проверка сохранения моделей

- 1) Убедиться, что при нажатии на кнопку "Dump models" созданные модели корректно сохраняются в JSON-файл.
- 2) Проверить структуру и содержимое сгенерированного JSON-файла на соответствие заданным параметрам.

6.2.1.3. Проверка генерации данных

- 1) Убедиться, что при нажатии на кнопку "Generate data" открывается диалоговое окно для указания параметров генерации.
- 2) Проверить, что генерируются данные с указанным количеством признаков и наблюдений.
- 3) Проверить, что сгенерированные данные сохраняются в файл "sample.csv" в корректном формате.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6.2.1.4. Проверка запуска анализа

- 1) Проверить, что при нажатии на кнопку "Run on models" запускается анализ на основе сохраненных моделей.
- 2) Убедиться, что результаты анализа визуализируются в виде графиков для каждой модели.

6.2.1.5. Проверка подсчета статистик

- 1) Проверить, что при нажатии на кнопку "Calculate statistics" запускается открывается диалоговое окно для ввода пути к CSV-файлу с данными.
- 2) Убедиться, что подсчитанные статистики вывелись корректно на экран.

6.2.2. Модуль статистических методов**6.2.2.1. Метод наименьших квадратов (LSM)**

- 1) Убедиться, что метод правильно вычисляет коэффициенты регрессии.
- 2) Проверить устойчивость метода при отсутствии аномальных наблюдений.

6.2.2.2. Метод наименьших абсолютных отклонений (LAD):

- 1) Проверить корректность вычисления коэффициентов модели.
- 2) Убедиться в устойчивости метода при наличии выбросов в данных.

6.2.2.3. Робастные методы регрессии (Huber, Tukey)

- 1) Проверить корректность реализации функций потерь и их градиентов.
- 2) Убедиться, что методы корректно работают при различных значениях параметра дельта.
- 3) Проверить сходимость методов при указанном количестве эпох и скорости обучения.

6.2.2.4. Проверка реализации метода Тейла-Сена (TheilSen)

- 1) Убедиться в корректности вычисления медианных наклонов.
- 2) Проверить устойчивость метода к наличию выбросов в данных.

6.2.3. Модуль машинного обучения

- 1) Проверить корректность работы алгоритма Isolation Forest.
- 2) Проверить корректность работы алгоритма DBSCAN.
- 3) Проверить корректность работы алгоритма KDE (Kernel Density Estimation).
- 4) Проверить корректность работы алгоритма KNN для обнаружения аномалий.
- 5) Убедиться, что алгоритмы корректно идентифицируют аномальные

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

наблюдения.

6.2.4. Модель визуализации

6.2.4.1. Проверка генерации графиков

- 1) Убедиться, что создаются корректные конфигурационные файлы для графиков.
- 2) Проверить, что графики корректно отображают зависимость ошибки от уровня шума.
- 3) Убедиться в правильном форматировании осей, заголовков и легенд на графиках.

6.2.4.2. Проверка сохранения результатов

- 1) Убедиться, что графики сохраняются в формате PNG с указанными именами.
- 2) Проверить качество и читаемость сохраненных изображений.
- 3) Убедиться, что метрики качества регрессии и обнаружения аномальных наблюдений сохраняются с указанным именем, содержимое файла не пусто, метрики принимают неотрицательные значения.

6.3. Проверка требований к надёжности

6.3.1. Проверка устойчивости программы к некорректным входным данным

- 1) Убедиться, что программа корректно обрабатывает отсутствующие файлы данных.
- 2) Проверить реакцию программы на некорректные параметры в JSON-файле.
- 3) Убедиться в корректной обработке ошибок при генерации данных с недопустимыми параметрами.

6.3.2. Проверка параллельной обработки данных

- 1) Убедиться, что программа эффективно использует доступные вычислительные ресурсы.
- 2) Проверить корректность работы механизма асинхронного выполнения экспериментов.

6.3.3. Проверка работы с памятью

- 1) Убедиться в отсутствии утечек памяти при длительной работе программы.
- 2) Проверить корректное освобождение ресурсов при завершении работы.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Произведя проверку функциональных требований в п. 6.2 и требований к надежности, можно сделать вывод, что программный комплекс удовлетворяет заявленным требованиям и обеспечивает надежную работу во всех тестовых сценариях.

СПИСОК ИСТОЧНИКОВ

- 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 19.603-78 Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 9) ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 10) ГОСТ 19.404-79 Программа и методика испытаний. Требования к содержанию и оформлений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 11) ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению. //

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

12) ГОСТ 19.504-79 Руководство программиста. Требования к содержанию и оформлению.

// Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

13) ГОСТ 19.401-78 Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

- 14) ГОСТ 19.505-79 Руководство оператора. Требования к содержанию и оформлению // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.11.04-01 ПМИ 03-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

[illegible]