

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ
Школа профессионального и академического образования

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

Директор ШПиАО
Д.В. Денисов
(подпись) (Ф.И.О.)
« » 2024 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ИССЛЕДОВАНИЕ МЕТОДОВ ОБРАБОТКИ СИГНАЛОВ
УЛЬТРАЗВУКОВЫХ РАСХОДОМЕРОВ С ИСПОЛЬЗОВАНИЕМ
АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Научный руководитель: Ронкин Михаил Владимирович
к.т.н., доцент


подпись

Нормоконтролер: Огуренко Егор Владимирович

подпись

Студент группы РИМ-22096: Гаврилин Пётр Александрович


подпись

Екатеринбург
2024

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования

«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий – РТФ
Школа профессионального и академического образования
Направление подготовки 09.04.01 Информатика и вычислительная техника
Образовательная программа 09.04.01/33.03 Инженерия машинного обучения

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студента Гаврилина Петра Александровича группы РИМ-220962
(фамилия, имя, отчество)

1. Тема выпускной квалификационной работы Исследование методов обработки
сигналов ультразвуковых расходомеров с использованием алгоритмов машинного обучения

Утверждена распоряжением по институту от «4» декабря 2023 г. № 33.02-05/298

2. Научный руководитель

Ронкин Михаил Владимирович, доцент, кандидат технических наук, доцент
(Ф.И.О., должность, ученая степень, ученое звание)

3. Исходные данные к работе результаты эксперимента (показания эталонного
расходомера, сигналы биения ПЭП), научные статьи с описанием эксперимента

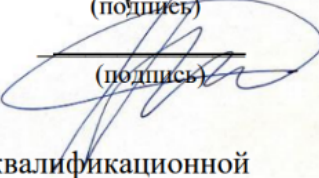
4. Перечень демонстрационных материалов презентация, jupyter-ноутбуки на github.com

5. Календарный план

№ п/п	Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1.	<i>1 раздел (глава)</i>	до 23.03.2024 г.	✓
2.	<i>2 раздел (глава)</i>	до 29.04.2024 г.	✓
3.	<i>3–4 раздел (глава)</i>	до 20.05.2024 г.	✓
4.	<i>ВКР в целом</i>	до 24.05.2024 г.	✓

Научный руководитель Ронкин Михаил Владимирович
Ф.И.О.

Студент задание принял к исполнению _____
дата


(подпись)

(подпись)

6. Допустить Гаврилина Петра Александровича к защите выпускной квалификационной работы в экзаменационной комиссии

Директор ШПиАО

(подпись)

Д.В. Денисов
Ф.И.О.

РЕФЕРАТ

Пояснительная записка 60 стр., 41 рис., 24 табл.

МАШИННОЕ ОБУЧЕНИЕ, НЕЙРОННАЯ СЕТЬ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ, УЛЬТРАЗВУКОВОЙ РАСХОДОМЕР.

В работе выполнено исследование методов обработки сигналов ультразвуковых расходомеров с использованием алгоритмов машинного обучения.

Проведён литературный обзор, охватывающий принципы применения алгоритмов машинного обучения в метрологии, практику применения машинного обучения для повышения точности ультразвуковых расходомеров, обзор методов машинного обучения и их применимости для поставленных в работе задач.

Выполнен анализ результатов эксперимента, в рамках которого получены данные расхода, измеренные эталонным расходомером, и соответствующие ему сигналы биения с датчиков ультразвукового расходомера. Разработана методология, определены метрики оценки точности и методы разработки моделей машинного обучения, в том числе применимые библиотеки языка программирования Python.

Осуществлена разработка и подбор гиперпараметров моделей машинного обучения. На основании анализа полученных метрик сделан вывод о целесообразности использования свёрточных нейронных сетей и линейной регрессии для калибровки ультразвуковых расходомеров.

СОДЕРЖАНИЕ

РЕФЕРАТ	3
СОДЕРЖАНИЕ	4
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	6
ВВЕДЕНИЕ	8
ГЛАВА 1. АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ ОБРАБОТКИ СИГНАЛОВ УЛЬТРАЗВУКОВЫХ РАСХОДОМЕРОВ И ПРАКТИКИ ПРИМЕНЕНИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В МЕТРОЛОГИИ.....	10
1.1 Методы обработки сигналов ультразвуковых расходомеров	10
1.1.1 Принцип работы ультразвукового расходомера	10
1.1.2 Погрешность измерения ультразвукового расходомера	11
1.2 Практика применения методов машинного обучения в метрологии	12
1.2.1 Принципы применения машинного обучения в метрологии	12
1.2.2 Использование машинного обучения в расходомерии	13
1.3 Алгоритмы обработки данных и машинного обучения.....	15
1.3.1 Сглаживание данных	15
1.3.2 Извлечение признаков.....	16
1.3.3 Линейная регрессия	16
1.3.4 К-ближайших соседей	17
1.3.5 Методы с использованием деревьев и ансамблей	18
1.3.6 Нейронные сети.....	18
1.4 Выводы из главы 1	20
ГЛАВА 2. МЕТОДОЛГИЯ ИССЛЕДОВАНИЯ МЕТОДОВ ОБРАБОТКИ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ	22

2.1 Программное обеспечение и сервисы, использованные в эксперименте	22
2.2 Схема получения экспериментальных данных.....	22
2.3 Обработка данных при исследовании.....	25
2.4 Метрики, применяемые в исследовании.	26
2.5 Алгоритмы линейной регрессии и k-ближайших соседей	27
2.6 Алгоритмы нейронных сетей.....	27
2.7 Выводы из главы 2	28
ГЛАВА 3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ БИЕНИЯ ПЬЕЗОЭЛЕКТРИЧЕСКИХ ПРЕОБРАЗОВАТЕЛЕЙ УЛЬТРАЗВУКОВЫХ РАСХОДОМЕРОВ	29
3.1 Результаты обработки данных.....	29
3.2 Исследование алгоритмов линейной регрессии и k-ближайших соседей	29
3.2.1 Результаты построения модели линейной регрессии	29
3.2.2 Результаты построения модели k-случайных соседей.....	35
3.3 Исследование нейронных сетей	39
3.3.1 Результаты построения полносвязной нейронной сети.....	39
3.3.2 Результаты построения свёрточной нейронной сети.....	44
3.3.3 Результаты построения рекуррентной нейронной сети.....	50
3.4 Выводы из главы 3	55
ЗАКЛЮЧЕНИЕ	56
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	57

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ПЭП – пьезоэлектрический преобразователь

MSE – среднеквадратичная ошибка

RMSE – корень среднеквадратичной ошибки

R² – коэффициент детерминации

Q – расход жидкости

D – внутренний диаметр трубопровода

Δt – разность времён распространения ультразвукового сигнала по и против потока

L_a – длина активной части акустического канала

C_0 – скорость ультразвука в неподвижной среде

α – угол между осями трубопровода и датчиков ПЭП

K_k – коэффициент коррекции, рассчитываемый эмпирически в процессе поверки расходомера

MA_t – значение скользящего среднего в точке t

n – количество значений исходной функции для расчёта скользящего среднего

ω_{t-i} – нормированный вес (t-i)-го значения исходной функции

p_{t-i} – значение исходной функции

y_i – предсказанное значение

$\tilde{y}_i$ – фактическое значение

$y_{\text{ср}}$ – среднее фактическое значение

n – размер выборки

Y – зависимая переменная линейной регрессии

X_i – независимые переменные линейной регрессии

b_i – коэффициенты линейной регрессии

ϵ – ошибка, характеризующая отклонение фактических данных от прогнозных

L1 – функция потерь при L1-регулязации

$L2$ – функция потерь при L1-регулязации

λ – гиперпараметр при L1- и L2-регуляризации

LSTM – рекуррентная нейронная сеть типа «долгая краткосрочная память»

GRU – рекуррентная нейронная сеть типа «управляемый рекуррентный блок»

ВВЕДЕНИЕ

Современное развитие вычислительной техники и повышение скорости обработки данных представляет дополнительные возможности по использованию машинного обучения и нейронных сетей для анализа и обработки сигналов датчиков измерительных приборов.

С учётом того, что при решении метрологических задач имеется возможность создания больших обучающих выборок, использование способности нейронных сетей решать задачи большой размерности является целесообразным [1, с. 81]. Применение соответствующих алгоритмов в промышленности может обеспечить повышение точности измерений, увеличение скорости калибровки и градуировки, а также возможность дообучения разработанных моделей на этапе эксплуатации [2, с. 64].

Повышение точности времяпролётных ультразвуковых расходомеров является актуальной проблемой, так как существующие методы обработки сигналов пьезоэлектрических преобразователей (далее – ПЭП) предполагают сложные алгоритмы расчёта и использование эмпирических коэффициентов, в результате чего точность измерения падает при отклонении от величины расхода, при которой производилась калибровка прибора [3, с. 363].

Сложность градуировки ультразвуковых расходомеров приводит к необходимости поиска новых способов обработки сигналов ультразвуковых расходомеров без расчёта поправочных гидравлических коэффициентов. Данная проблема может быть решена с помощью машинного обучения.

В настоящее время успешно ведутся разработки применения методов машинного обучения и нейронных сетей для обработки сигналов, имеющих характеристики временных рядов с периодическими колебаниями, в том числе: при классификации радиосигналов [4][5], в приборах неразрушающего контроля, использующих ультразвук и акустическую эмиссию [6], в вибродиагностике [7], в медицинской диагностике [8] [9], а также непосредственно в расходомерии жидкостей и газов [10] [11] [12] [13].

Объектом исследования в данной работе являются накладные ультразвуковые расходомеры со сложномодулированными сигналами.

Предмет исследования: методы вычисления значения расхода по сигналам биения ультразвукового расходомера.

Целью работы является исследование методов обработки сигналов ультразвуковых расходомеров с использованием алгоритмов машинного обучения, направленных на повышение точности за счёт исключения из расчёта расхода гидравлического корректирующего коэффициента.

Для достижения целей исследования были поставлены и выполнены следующие задачи:

1. Анализ существующих методов обработки сигналов датчиков ультразвуковых расходомеров и практики применения машинного обучения в метрологии.

2. Разработка методологии проведения исследования методов цифровой обработки сигналов ультразвуковых расходомеров.

3. Исследование алгоритмов машинного обучения, применяемых в качестве методов цифровой обработки сигналов ультразвуковых расходомеров.

В рамках работы разработаны и обучены модели машинного обучения, применимые для использования в программном обеспечении ультразвуковых расходомеров, которые решают проблемы возникновения дополнительной погрешности из-за использования эмпирического гидравлического коэффициента коррекции.

Результаты исследования имеют научную новизну: получило дальнейшее развитие применение моделей машинного обучения для целей повышения точности измерений ультразвуковых расходомеров жидкости.

Результаты исследования опубликованы в сборниках тезисов Всероссийской конференции «ИНТЕР – Информационные технологии и радиоэлектроника» и XX Всероссийской конференции «Информационные технологии в современном мире – 2024».

ГЛАВА 1. АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ ОБРАБОТКИ СИГНАЛОВ УЛЬТРАЗВУКОВЫХ РАСХОДОМЕРОВ И ПРАКТИКИ ПРИМЕНЕНИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В МЕТРОЛОГИИ.

1.1 Методы обработки сигналов ультразвуковых расходомеров

1.1.1 Принцип работы ультразвукового расходомера

Принцип работы ультразвукового расходомера базируется на закономерности, согласно которой скорость ультразвука в движущейся среде равна сумме скорости ультразвука в неподвижной жидкости и проекции скорости потока на направление распространения ультразвука.

В ультразвуковом расходомере два ПЭП располагаются под углом к потоку жидкости и выступают поочерёдно излучателями и приёмниками, что позволяет вычислить время прохождения ультразвукового сигнала по направлению потока и против него. Разница этих времён пропорциональна скорости потока и, как следствие, расходу жидкости [14, с. 17].

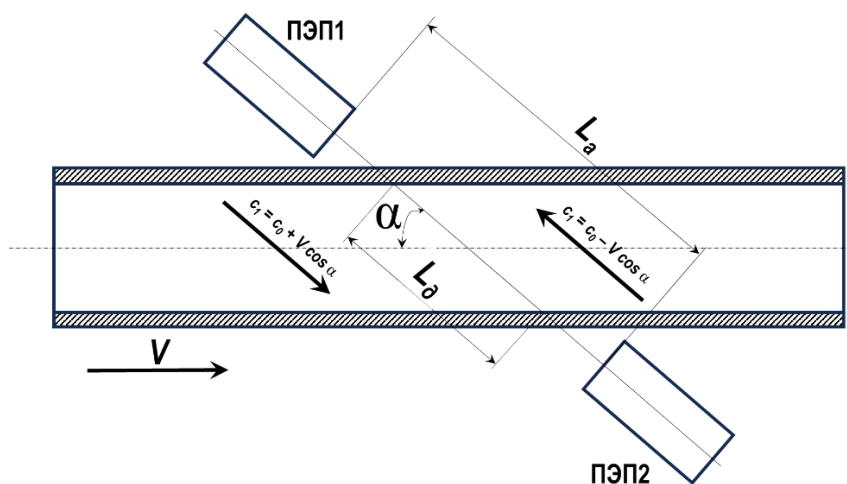


Рисунок 1.1 – Схема ультразвукового расходомера

Расход жидкости таким расходомером рассчитывается по формуле [14, с. 18]:

$$Q = \frac{\pi D^2 \cdot K_k}{4} * \frac{\Delta t \cdot C_0^2}{2 \cdot L_a \cdot \cos \alpha}, \quad (1.1)$$

где D – внутренний диаметр трубопровода;

Δt – разность времён распространения ультразвукового сигнала по и против потока;

L_a – длина активной части акустического канала;

C_0 – скорость ультразвука в неподвижной среде;

α – угол между осями трубопровода и датчиков ПЭП;

K_k – коэффициент коррекции, рассчитываемый эмпирически в процессе поверки расходомера.

Принцип действия ПЭП основан на прямом и обратном пьезоэлектрическом эффекте, суть которого заключается в возникновении электрических зарядов при приложении к пьезоэлектрикам механической силы, и наоборот, возникновении механических колебаний при подключении пьезоэлемента к напряжению электрического тока [15, с. 13].

Основным методом для определения времени прохождения ультразвука от ПЭП-излучателя до ПЭП-приёмника (времени задержки) является импульсно-пороговый метод [16, с. 2331].

Более высокую точность измерений даёт использование линейно-частотного модулирования, в рамках которого задача определения времени задержки сводится к измерению частоты сигналов биения. При этом сигнал биения – это низкочастотная составляющая произведения излученного и опорного сигналов гетеродина [17, с. 2326].

Существуют различные модификации ультразвуковых расходомеров, различающиеся по количеству и способу установки ПЭП.

1.1.2 Погрешность измерения ультразвукового расходомера

На погрешность ультразвуковых времяпролётных расходомеров могут воздействовать различные факторы.

В результате разности скорости средней скорости по пути перемещения акустических колебаний и средней скорости потока неправильный учёт профиля скоростей. Для снижения воздействия данного фактора применяется корректирующий коэффициент K_k . Однако при этом необходимо учитывать, что наибольшую точность прибор будет показывать на расходе, при котором выполнялась калибровка. При отклонении от калибровочного расхода погрешность расходомера возрастает [3, с. 363].

Также, на погрешность ультразвуковых расходомеров влияют следующие дополнительные факторы:

- температура и давление измеряемой среды и те параметры, на которые они влияют, в том числе плотность среды и геометрические размеры трубопровода и ПЭП;
- неравномерность плотности среды и наличие примесей;
- наличие паразитарных шумов, возникающих за счёт прохождения ультразвука от датчика к датчику по материалу трубы и переотражений сигнала от стенок трубы;
- наличие паразитарных шумов от расположенного на трубопроводе оборудования (насосы, компрессоры, подшипники качения и т.д.) [18, с. 114].

Для снижения влияния вышеуказанных факторов разрабатываются различные алгоритмы обработки сигналов, направленные на фильтрацию шумов и выделение полезного сигнала.

1.2 Практика применения методов машинного обучения в метрологии

1.2.1 Принципы применения машинного обучения в метрологии

Применение методов машинного обучения для цифровой обработки сигналов датчиков измерительных приборов имеет значительные перспективы, так как может обеспечить более высокую точность измерений за счёт нивелирования проблем, связанных со сложностью математического описания преобразования сигналов датчиков в измеряемую величину [19, с. 2].

При этом при выборе машинного обучения и нейронных сетей в качестве алгоритма цифровой обработки сигналов датчиков измерительного прибора необходимо обязательное соблюдение трёх базовых принципов [10, с. 30]:

целесообразность: при наличии простых и досконально изученных функций преобразования сигналов датчиков прибора в измеряемую величину применение машинного обучения и нейронных сетей нерационально, так как не приведёт к повышению точности измерений, но при этом вызовет рост стоимости оборудования из-за необходимости обеспечить дополнительную вычислительную мощность;

доступность данных: для формирования обучающих и тестовых выборок требуется наличие возможности получения экспериментальных данных необходимой точности, для чего необходимы испытательные стенды с эталонными измерительными приборами, обеспечивающие условия и режимы работы, которые соответствуют реальным условиям дальнейшего использования измерительного оборудования;

релевантность: необходимо наличие взаимосвязи между сигналами датчиков и измеряемой величиной, так как нейронная сеть способна просчитывать неочевидные или неизвестные зависимости, но только в том случае, если они существуют.

Кроме того, алгоритмы машинного обучения имеют возможность точной градуировки для конкретных условий использования измерительных приборов, а также возможность дообучения на этапе их эксплуатации.

1.2.2 Использование машинного обучения в расходометрии

В расходометрии был сертифицирован в качестве средства измерения и нашёл коммерческое применение при измерениях дебита нефтяных скважин многофазный кориолисовый расходомер, который использует для цифровой обработки сигналов искусственную нейронную сеть.

В данном случае была решена проблема роста погрешности кориолисного расходомера при росте объёмной доли попутного нефтяного газа в добываемом сырье. На вход нейронной сети в данном случае подаются показания расходомера по расходу и плотности, показания влагомера по обводнённости, показания многопараметрического датчика давления, температурного датчика и отклонение измеренной плотности от расчётной плотности жидкой фазы. На выходе – скорректированный показатель массового расхода [10, с. 30].

Для ультразвуковых расходомеров проводились исследования по снижению ошибки измерения, связанной с применением коэффициента коррекции. Авторами предложено корректирующий гидравлический коэффициент K_k сделать переменным в зависимости от входящих параметров. В простейшем случае была построена одномерная линейная регрессия, построенная с использованием метода наименьших квадратов на сопоставлении фактической скорости потока и скорости, измеренной расходомером. После этого линейная регрессия была дополнена дополнительным признаком: числом Рейнольдса, а также проведена классификация ошибок наивным байесовским методом, и значительные ошибки из данных были удалены [11, с. 15].

Более сложная задача решалась с помощью полносвязной нейронной сети, которая имела 10 входов, 8 нейронов на первом слое, 2 нейрона на втором слое и 1 выход. Входными данными при этом были обработанные сигналы с ПЭП. [13, с. 60].

Также, есть опыт применения полносвязной нейронной сети из двух слоёв для исследования зависимости расхода от таких входных параметров, как диаметр трубопровода, плотность, температура жидкости и напряжение на преобразователе частоты с усилителем [12, с. 192]. Это исследование выявило возможность применения нейронной сети для настройки расходомеров таким образом, чтобы смена параметров трубопровода и транспортируемой среды не приводило к необходимости новой калибровки прибора. При этом данный

подход не решает проблему неравномерности скоростей в потоке и применения корректирующего коэффициента.

Необходимо констатировать, что на текущий момент в ультразвуковой расходомерии не получило должного распространения использования всего спектра инструментов программной обработки данных и машинного обучения.

1.3 Алгоритмы обработки данных и машинного обучения

1.3.1 Сглаживание данных

При применении машинного обучения для обеспечения точности и производительности моделей необходимо производить предварительную обработку данных и извлечение признаков.

В случае, если колебания значений во временном ряде существенно влияют на точность предсказания и фактически могут являться шумами и помехами необходимо применять математические методы, сглаживающие данные.

Наиболее часто используемым при этом является метод скользящего среднего [20, с. 50]. Скользящее среднее является дискретным случаем свёртки двух функций и рассчитывается по формуле:

$$MA_t = \sum_{i=0}^{n-1} \omega_{t-i} \cdot p_{t-i}, \quad (1.2)$$

где MA_t – значение скользящего среднего в точке t ;

n – количество значений исходной функции для расчёта скользящего среднего;

ω_{t-i} – нормированный вес $(t-i)$ -го значения исходной функции;

p_{t-i} – значение исходной функции.

Данный метод реализован в библиотеке Pandas языка Python функцией `rolling()`, в котором параметром `win_type` задаётся тип сглаживания: среднеарифметическое, гауссианное, экспоненциальное, медианное и т.д.

Также для сглаживания данных эффективно использование фильтра Савицкого-Голея, частным случаем которого является метод скользящего арифметического среднего [21, с. 73]. Отличие заключается в том, что в фильтре Савицкого-Голея с использованием метода наименьших квадратов строится аппроксимирующий полином необходимого порядка. Метод, несмотря на математическую сложность, имеет низкую вычислительную мощность и зарекомендовал себя на практике [22, с. 52]. Данный метод реализован в библиотеке `scipy` функцией `signal.savgol_filter()`.

1.3.2 Извлечение признаков

Большое количество признаков может привести к таким негативным последствиям, как низкая производительность и, в отдельных случаях, снижение точности модели машинного обучения. В связи с этим разработаны различные методы снижения размерности и извлечения признаков.

Самым простым из таких методов является метод анализа главных компонент PCA (principal component analysis), суть которого заключается в построении линейных проекций для преобразования данных в новое пространство, построенного путём поиска корреляционных связей между признаками [23, с. 129].

Другим методом является T-распределённое стохастическое вложение соседей tSNE (T-distributed stochastic neighbor embedding), основанный в отличие от PCA на нелинейных закономерностях. Главный принцип метода заключается в том, чтобы сходные объекты сохраняли свою похожесть в новом пространстве [24, с. 98].

Для извлечения признаков временных рядов в Python разработана библиотека Time Series Feature Extraction Library (TSFEL) [25].

1.3.3 Линейная регрессия

Линейная регрессия является самым простым и производительным методом машинного обучения.

Линейная регрессия исходит из допущения, что зависимая переменная Y является линейной функцией независимых переменных X_i , в которой модель подбирает коэффициенты b_i , минимизируя функцию потерь с помощью метода наименьших квадратов.

$$Y = b_0 + b_1 \times X_1 + b_2 \times X_2 + \dots + b_n \times X_n + \varepsilon, \quad (1.3)$$

В библиотеках для python реализована функцией `sklearn.linear_model.LinearRegression`.

Для решения проблем с переобучением модели, когда на обучающих данных модель показывает хорошие результаты, а на реальных даёт большую ошибку, используют методы регуляризации [26, с. 1755].

L1-регуляризация (`sklearn.linear_model.Lasso`) заключается в добавлении к функции потерь штрафа, равного сумме модулей коэффициентов модели умноженных, на гиперпараметр λ . L1-регуляризация способна убирать неинформативные признаки из модели за счёт обнуления их весов, что позволяет уменьшить сложность модели улучшить обобщающую способность.

$$L1 = \sum (y_i - \bar{y}_i)^2 + \lambda \times \sum |b_i| \quad (1.4)$$

При L2-регуляризации (`sklearn.linear_model.Ridge`) в штрафе вместо модулей весов складываются их квадраты. Это позволяет более гладко и непрерывно штрафовать большие значения весов, не зануляя их полностью.

$$L2 = \sum (y_i - \bar{y}_i)^2 + \lambda \times \sum b_i^2 \quad (1.5)$$

Комбинация L1- и L2-регуляризации реализована в функции `sklearn.linear_model.ElasticNet`.

1.3.4 К-ближайших соседей

Другим алгоритмом решения регрессионной задачи является метод k ближайших соседей, в котором значение метки вычисляется на основании среднего значения меток ближайших соседей. При этом гиперпараметром является либо количество ближайших соседей k

(*sklearn.neighbors.KNeighborsRegressor*), либо радиус r , в котором они должны находиться (*sklearn.neighbors.RadiusNeighborsRegressor*).

К недостаткам данного метода относится высокая стоимость вычисления и плохая работа с категориальными признаками.

Метод k -ближайших соседей чаще применяется в задачах классификации, чем в задачах регрессии.

1.3.5 Методы с использованием деревьев и ансамблей

Также, как и метод k -ближайших соседей, деревья решений чаще применяются в задачах классификации. При применении их (*sklearn.tree.DecisionTreeRegressor*) к задачам регрессии мы на листьях имеем не класс, а конкретное значение метки.

Также, используется для задач регрессии алгоритм градиентного бустинга (*sklearn.ensemble.GradientBoostingRegressor*), в котором предсказание строится в виде ансамбля слабых предсказывающих моделей, в качестве которых в основном используются деревья решений.

Для градиентного бустинга кроме *sklearn* возможно использование библиотек *XGBoost*, *LightGBM*, *CatBoost*, которые при определённых задачах имеют преимущества по скорости обработки данных и точности.

Методы, использующие деревья решений, не подходят для решаемой задачи обработки сигналов ультразвуковых расходомеров, так как в обучающей выборке будет конечное дискретное количество расходов, на которых будет выполняться градуировка.

1.3.6 Нейронные сети

Нейронные сети являются моделями глубокого машинного обучения, основным признаком которых является наличие слоёв, на каждом из которых происходит преобразование (очищение) информации, необходимое для решения задачи [27, с. 34]. Они находят широкое применение для совершенно различных задач, включая распознавание и генерацию изображений,

транскрибацию речи, обработку текстов, создание чат-ботов и аналитических помощников, автопилоты, обработки сигналов различного вида и т.д.

Общий принцип работы полносвязной нейронной сети заключается в преобразовании входных данных в каждом нейроне путём сложения произведений входных значений на соответствующие веса и применения к получившейся сумме функции активации, после чего вычисленные значения передаются на следующий слой в качестве входных данных.

Обучение нейронной сети заключается в подборе вышеуказанных весов таким образом, чтоб выходной слой соответствовал корректным ожидаемым значениям, то есть задачей в данном случае является минимизация ошибки.

Существует различные виды нейронных сетей для обучения с учителем, которые отличаются друг от друга принципами обработки информации в нейронах, из которых стоит выделить рекуррентные и свёрточные нейронные сети.

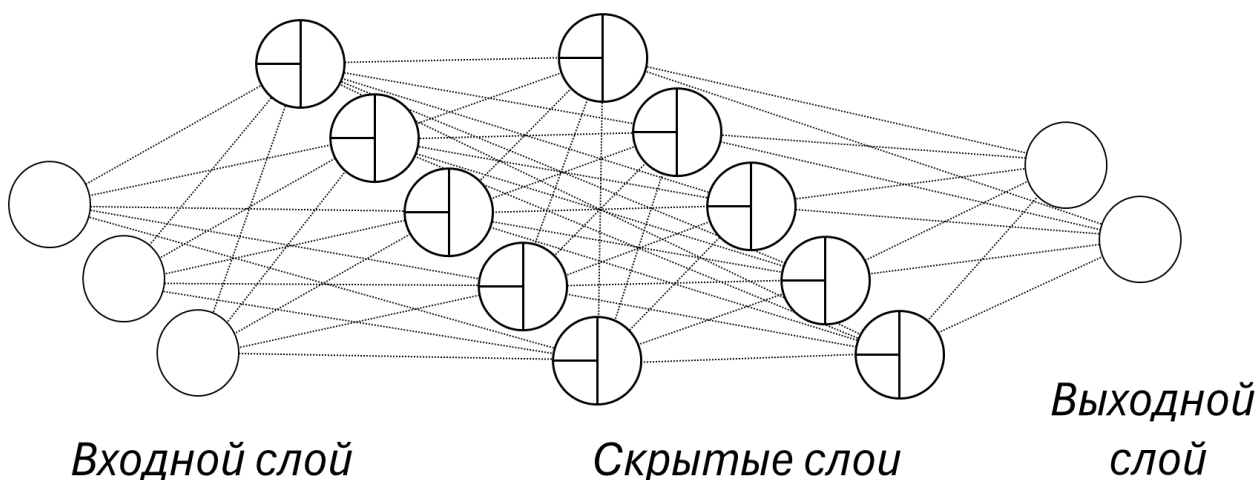


Рисунок 1.2 – Общая схема полносвязной нейронной сети

Особенностью рекуррентных нейронных сетей является наличие связи между соседними элементами входных данных слоя. Существуют различные виды рекуррентных сетей, основными из которых являются LSTM (долгая краткосрочная память) и GRU (управляемый рекуррентный блок).

Рекуррентные нейронные сети за счёт эффекта памяти активно используются в задачах распознавания текста, то есть в моделировании временных взаимосвязанных последовательностей [28, с. 315]. Таким образом, есть основания сделать предположение применимости рекуррентных нейронных сетей для целей обработки сигналов датчиков ультразвуковых расходомеров.

В свёрточных нейронных сетях ключевой операцией является операция свёртки. Операцией свёртки называется применение градиентного фильтра к части массива данных. Применение операции свёртки ко всему массиву данных с определённым шагом образует свёрточный слой.

Свёрточные нейронные сети успешно применялись для распознавания типа модуляций радиосигналов [5, с. 14], что также свидетельствует об их применимости для поставленных задач.

Для построения нейронных сетей возможно использование различных библиотек, в том числе PyTorch, TensorFlow и входящий в него Keras, Theano и другие. В библиотеке TSFEL также реализована возможность построение нейронных сетей для временных рядов.

При построении нейронной сети подбирают следующие гиперпараметры:

- количество слоёв;
- количество нейронов в каждом слое;
- функцию активации для каждого слоя;
- функцию потерь;
- оптимизатор;
- иные гиперпараметры.

1.4 Выводы из главы 1

Повышение точности ультразвуковых расходомеров является актуальной темой, так как данный вид измерительных приборов используется как в коммерческом, так и технологическом учёте. Они удобны в

обслуживании, не создают дополнительного сопротивления в трубопроводе и доступны по стоимости.

Однако существует проблема, связанная с ростом погрешности измерения при отклонении измеряемого расхода от значения, при котором производилась калибровка прибора и рассчитывался корректирующий гидравлический коэффициент. Это вызвано неравномерностью скоростей в профиле потока жидкостей.

Существующие методы цифровой обработки сигналов ультразвуковых расходомеров не решают данную проблему, так как в большей степени направлены на повышение точности измерения за счёт более корректного определения разности времён прохождения ультразвука по и против потока.

Описанные в литературе способы повышения точности ультразвуковых расходомеров с помощью алгоритмов машинного обучения сводятся к преобразованию корректировочного коэффициента K_k в переменную, зависящую от нескорректированных показаний расходомера и, иногда, других дополнительных параметров.

При этом ультразвуковые расходомеры полностью соответствуют базовым принципам применения машинного обучения в метрологии.

Таким образом, существует актуальная задача по разработке методологии исследования и апробированию моделей машинного обучения и нейронных сетей для пересчёта электрических сигналов пьезоэлектрических преобразователей ультразвуковых расходомеров в показания расхода жидкости или газа.

ГЛАВА 2. МЕТОДОЛГИЯ ИССЛЕДОВАНИЯ МЕТОДОВ ОБРАБОТКИ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

2.1 Программное обеспечение и сервисы, использованные в эксперименте

Для проведения исследования использован язык программирования Python и его библиотеки sys, os, numpy, pandas, scipy, sklearn, matplotlib, bottleneck, random, tsfel, tensorflow, plot_keras_history.

Расчёты выполнены в блокнотах Jupiter на сервисе colab.research.google.com.

Результаты работы расположены в открытом доступе по адресу https://github.com/PetrGavrilin/VKR_URFU.

2.2 Схема получения экспериментальных данных

Для проведения моделирования были использованы результаты исследования времяпролётного ультразвукового расходомера со сложномодулированными сигналами [29, с. 6031].

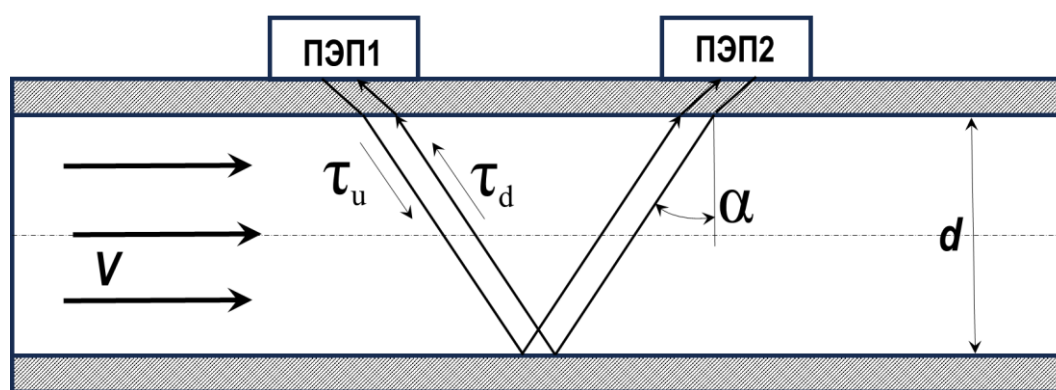


Рисунок 2.1 – Схема прохождения сигнала в эксперименте

На испытательном стенде экспериментально были получены для медной трубы диаметром 10 мм с толщиной стенки 2 мм следующие данные:

- расход воды в м³/ч., измеренный эталонным расходомером;

- значение напряжения электрического сигнала биения для каждого ПЭП в мВ;
- дополнительные параметры эксперимента (температура, разница времён прохождения ультразвукового сигнала, измеренная классическим способом, и т.д.)

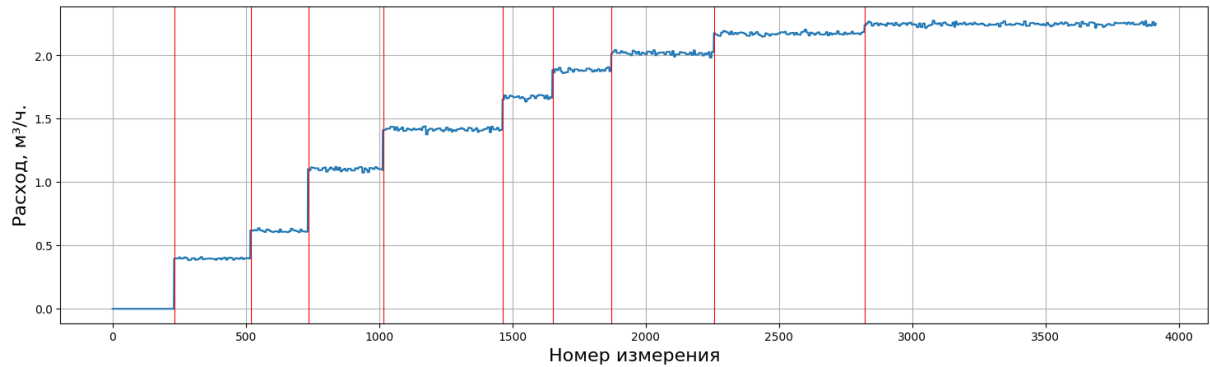


Рисунок 2.2 – Показания эталонного расходомера

В датасете 3 915 измерений. Выборка, соответствующая одному периоду линейно-частотной модуляции, имеет объем из 512 временных точек для каждого преобразователя.

На графике разности времён (рисунок 2.3) обращает на себя внимание отсутствие точной зависимости расхода от разности времён прохождения ультразвука в соответствии с формулой 1.1. То есть имеет место дополнительная погрешность измерения, связанная с учётом профиля скоростей, о которой упоминалось в главе 1.

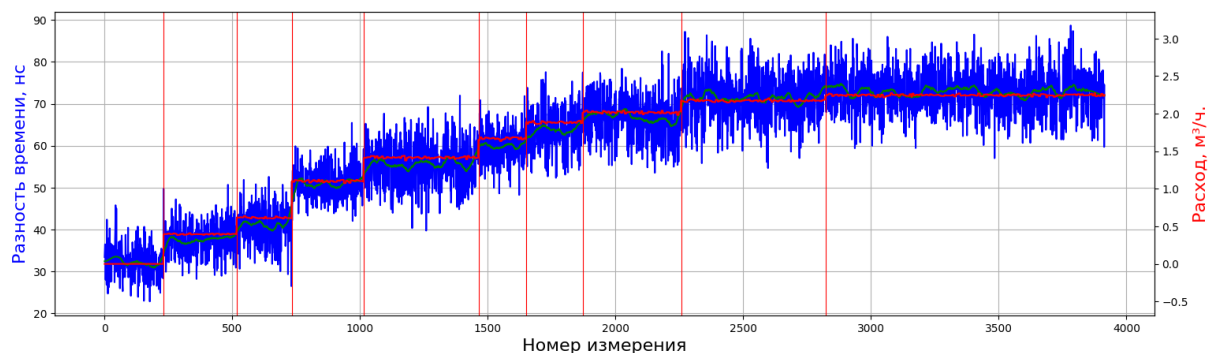


Рисунок 2.3 – Разность времён прохождения сигналов, определённая классическим методом

В части изменения температуры в течение проведения эксперимента, в связи с незначительными колебаниями показателя данным фактором предлагается пренебречь.

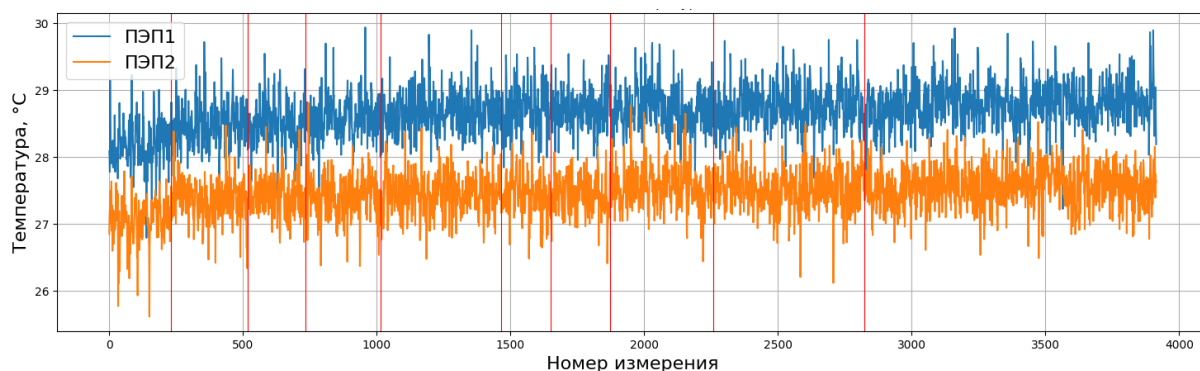


Рисунок 2.4 – Показатели температуры при проведении эксперимента

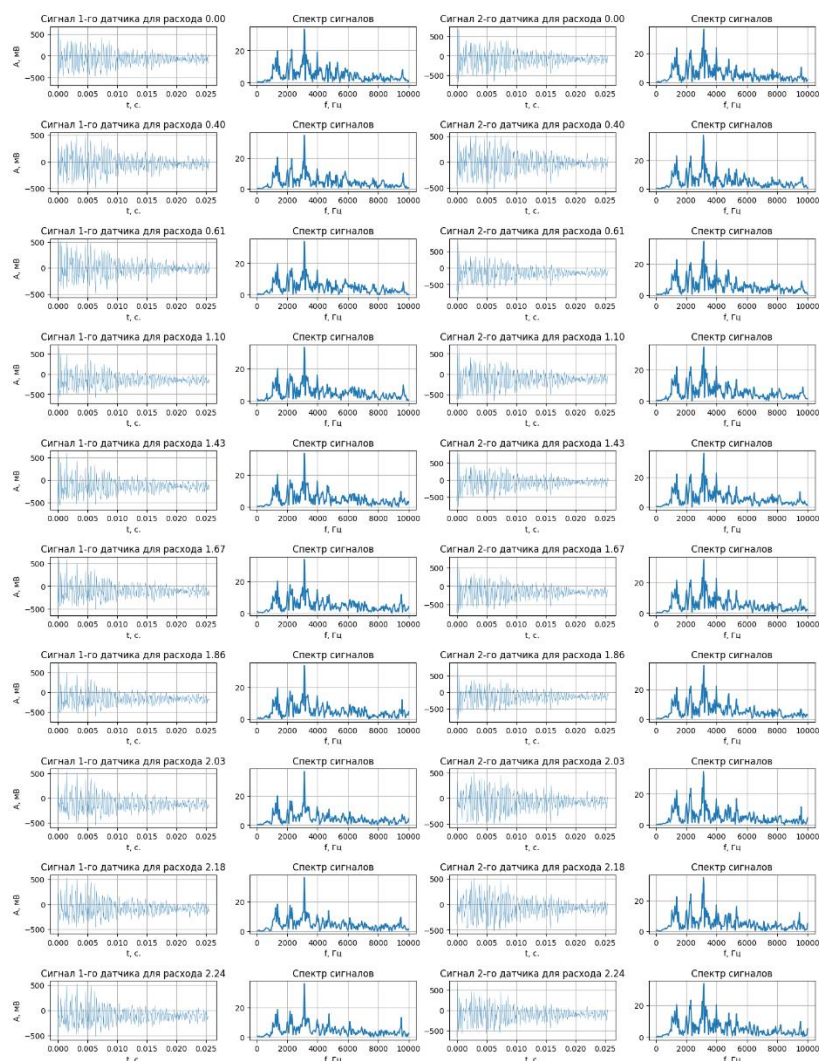


Рисунок 2.5 – Примеры сигналов биения для различных расходов и их спектры

На спектрах сигналов биения, полученных с помощью преобразования Фурье, (рисунок 2.5) кроме полезного сигнала видны дополнительные пики, что говорит о том, что сигнал достаточно сильно зашумлён из-за различных факторов, связанных с переотражением сигнала от стенок трубы, прохождением сигнала по стенке трубы напрямую и иными причинами.

Данные эксперимента не имеют пропусков, аномальных значений и пригодны для дальнейшей обработки и построения моделей машинного обучения.

2.3 Обработка данных при исследовании

Для сигналов ПЭП ультразвукового расходомера первичными данными будут показания напряжения электрического поля, возникающего при получении акустического сигнала. Такие данные в теории машинного обучения подпадают под определение временного ряда. В связи с этим появляется возможность использования всех существующих методов обработки и преобразования данных временных рядов.

Для изучения влияния шумов на точность построенных моделей машинного обучения в дальнейшем изучались два набора данных: исходный и сглаженный фильтром Савицкого-Голея с использованием функции `signal.savgol_filter()` библиотеки `scipy`, показавшего большую эффективность, чем метод скользящего среднего. Метод скользящего среднего применён только для линейной регрессии, где и была выявлена его меньшая эффективность.

Дальнейшая обработка обоих датасетов для исследования линейной регрессии и метода k-ближайших соседей состояла в:

- извлечении признаков с использованием функции `time_series_features_extractor()` библиотеки TSFEL;
- удалении сильно коррелированных признаков, выявленных функцией `correlated_features()` библиотеки TSFEL;

- удалении признаков с низкой дисперсией функцией с помощью класса `feature_selection.VarianceThreshold` библиотеки `sklearn`.

- стандартизации данных с использованием класса `preprocessing.StandardScaler` библиотеки `sklearn`.

Для проведения исследования нейронных сетей извлечение признаков не применялось. При обучении моделей к данным применялась пакетная нормализация.

С учётом того, что для корректности измерений необходимо обеспечивать стационарный режим расхода, возникает риск, что из-за того, что при градуировке невозможно учесть все режимы расходов, на значениях, неучтённых при градуировке, будут некорректные измерения.

Для оценки вышеуказанного предположения было выполнено дополнительное моделирование, в котором роль тестовой выборки выполнял один из режимов расхода – 1,42 м³/ч. (с 1015 по 1464 измерение).

2.4 Метрики, применяемые в исследовании.

С учётом того, что среднеквадратичная погрешность измерения в метрологии является классической [30, с. 5], для оценки точности работы моделей машинного обучения использовались следующие метрики:

- среднеквадратичная ошибка MSE и корень из неё RMSE;
- коэффициент детерминации R2.

Среднеквадратичная ошибка рассчитывается по формуле:

$$MSE = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}, \quad (2.1)$$

где y_i – предсказанное значение;

$\bar{y}_i$ – фактическое значение;

n – размер выборки.

Коэффициент детерминации рассчитывается по формуле:

$$R2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - y_{cp})^2}, \quad (2.1)$$

где y_i – предсказанное значение;

$\bar{y}_i$ – фактическое значение;

$y_{\text{ср}}$ – среднее фактическое значение;

n – размер выборки.

Для расчёта метрик использованы функции `metrics.mean_squared_error()` и `metrics.r2_score()` библиотеки `sklearn`.

Коэффициент детерминации не применялся к моделированию, где роль тестовой выборки выполнял режим расхода 1,42 м³/ч., так как из-за неучёта других режимов значение получается некорректным.

2.5 Алгоритмы линейной регрессии и k-ближайших соседей

Для линейной регрессии с и без регуляризацией использованы следующие функции подбиблиотеки `sklearn.linear_model`: `LinearRegression`, `Ridge`, `RidgeCV`, `Lasso`, `ElasticNetCV`, `SGDRegressor`.

Для каждой модели определены метрики, и модель с лучшими метриками использовалась для итогового моделирования и построения графиков.

Для метода k-ближайших соседей использована функция библиотеки `sklearn neighbors.KNeighborsRegressor()`. Гиперпараметры количества `n_neighbors` и весов `weights` подбирались вручную.

2.6 Алгоритмы нейронных сетей

Модели нейронных написаны с использованием библиотеки `tensorflow.keras`.

В результате подбора в качестве оптимизатора при обучении модели нейронной сети использован алгоритм адаптивной оценки импульса `Adam`, как показавший лучшие результаты. В качестве функции потерь использовалась среднеквадратичная ошибка (MSE).

В качестве функций активации использовались линейные функции `relu` и `relu6` и гиперболический тангенс `tanh`.

В качестве предобработки использовалась пакетная нормализация `BatchNormalization()`.

Полносвязная нейронная сеть была построена с использованием слоёв `Dense`. Оптимальное количество слоёв, количество нейронов и функции активации определены методом подбора.

Свёрточная нейронная модель построена на принципах, использованных в широко известной свёрточной нейронной сети VGG-16 [25, с. 207], используемой для распознавания изображений. Она состоит из стеков, в каждый из которых входит два свёрточных слоя `Conv1D()` и один слой пулинга `MaxPooling1D()`. Также для борьбы с переобучением использован слой регуляризации `Dropout()`, но в целом модели показывали устойчивость к переобучению, в связи с чем регуляризация применялась в незначительных объёмах. После свёрточных стеков для выпрямления данных в вектор применён слой `Flatten()`, после которого модель завершалась двумя полносвязными слоями `Dense()`.

Для построения рекуррентной нейронной сети был использован тип «управляемый рекуррентный блок» (Gated Recurrent Unit – GRU), как более производительный с точки зрения вычислительных мощностей. Количество слоёв и нейронов определялось подбором.

2.7 Выводы из главы 2

В рамках главы 2 выполнен анализ данных, полученных в результате эксперимента. Для построения в дальнейшем моделей использованы показания эталонного расходомера в метрах кубических в час и сигналы биения ПЭП в микровольтах.

Определена методология исследования методов цифровой обработки сигналов биения ультразвуковых расходомеров, в том числе выбраны метрики для оценки точности, алгоритмы предобработки данных и алгоритмы машинного обучения.

ГЛАВА 3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ БИЕНИЯ ПЬЕЗОЭЛЕКТРИЧЕСКИХ ПРЕОБРАЗОВАТЕЛЕЙ УЛЬТРАЗВУКОВЫХ РАСХОДОМЕРОВ

3.1 Результаты обработки данных

В результате извлечения признаков для данных изначального датасета без сглаживания осталось 616 из 1 024 признаков.

Для датасета, к которому применили фильтр Савицкого-Голея, осталось 508 признаков.

3.2 Исследование алгоритмов линейной регрессии и k-ближайших соседей

3.2.1 Результаты построения модели линейной регрессии

При подборе гиперпараметров линейной регрессии были опробованы различные варианты построения модели, в том числе с различными вариантами регуляризации. В итоге выявлено, что применение регуляризации существенно не влияет на повышение точности моделей для исследуемого набора данных.

Таблица 3.1 - Метрики регрессионных моделей для тестовых выборок данных без применения сглаживания

Метрика	Linear Regression()	Lasso()	Ridge()	ElasticNetCV()	RidgeCV()	SGD Regressor()
MSE	0.00120	0.51611	0.00120	0.00120	0.00120	0.18826
RMSE	0.03463	0.71841	0.03461	0.03455	0.03459	0.43389
R2	0.99768	-	0.99768	0.99769	0.99768	0.63521

В результатах моделирования, приведённых в таблицах 3.1 и 3.2, обращает на себя внимание одинаковая точность измерения для различных режимов расхода жидкости в трубопроводе (рисунок 3.1), что является

решением важной проблемы в ультразвуковой расходомерии, когда значение расхода сильно влияет на погрешность его измерения.

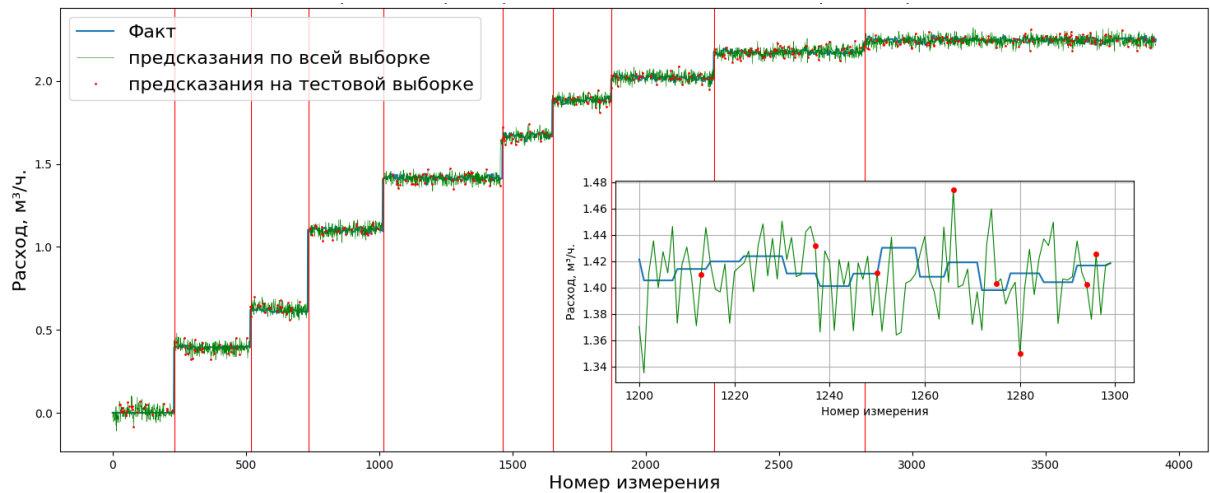


Рисунок 3.1 – Результаты предсказания модели RidgeCV для данных без применения сглаживания

Таблица 3.2 - Метрики модели RidgeCV для данных без применения сглаживания

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00120	0.00082	0.00086
RMSE	0.03459	0.02861	0.02926
R2	0.99768	0.99842	0.99834

Также, было выполнено моделирование, в котором роль тестовой выборки выполнял один из режимов расхода – 1,42 м³/ч. (с 1015 по 1464 измерение).

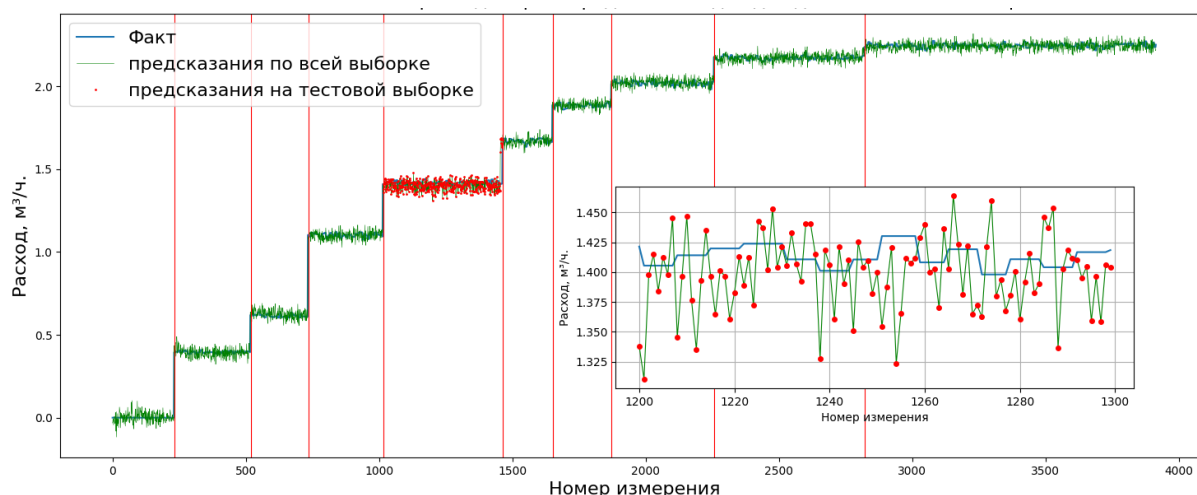


Рисунок 3.2 – Результаты предсказания модели Ridge для данных без применения сглаживания, где тестовой выборкой является один режим расхода

В результате моделирования, представленных в таблице 3.3, выявлено, что MSE увеличилась с 0.0012 до 0.0023, но при этом среднее значение расхода в режиме отличается от эталонного на 0.01 м³/ч. То есть ошибка увеличивается за счёт разброса значений вокруг корректного, а не за счёт смещения среднего арифметического.

Таблица 3.3 - Метрики модели Ridge для данных без применения сглаживания для данных без применения сглаживания, где тестовой выборкой является один режим расхода

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00230	0.00071	0.00089
RMSE	0.04796	0.02669	0.02991
R2	-	0.99877	0.99827

Учитывая характер колебания результатов работы модели, есть основания предполагать, что на их дисперсию влияют различные помехи, указанные в начале данной главы.

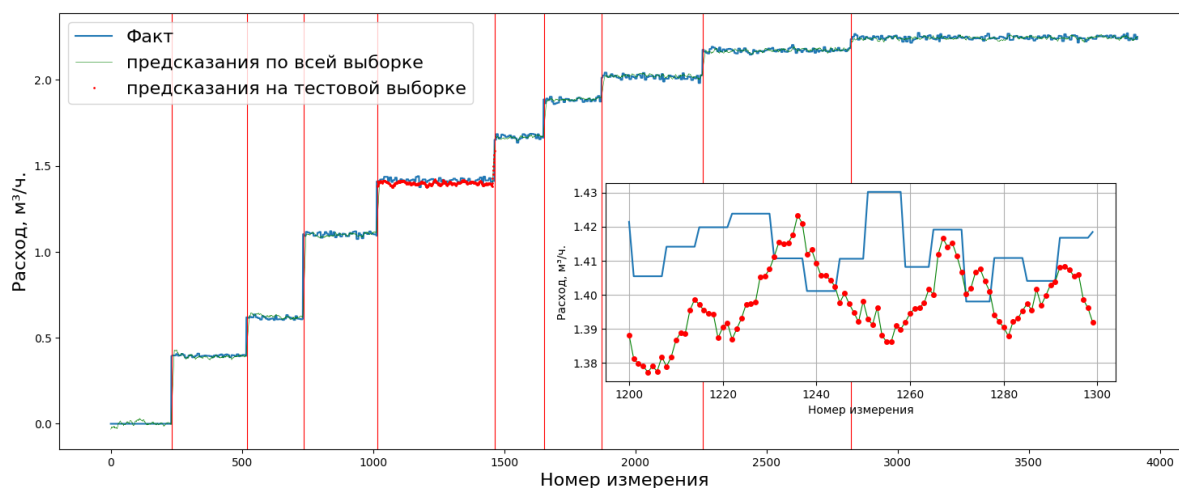


Рисунок 3.3 – Применение сглаживания к предсказаниям модели

В целях повышения точности измерения был применён метод скользящего среднего к результатам работы модели (рисунок 3.3), результаты моделирования после применения которого представлены в таблице 3.4.

Таблица 3.4 - Метрики модели Ridge после применения к результатам её работы метода скользящего среднего

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00063	0.00095	0.00075
RMSE	0.02502	0.03077	0.02746
R2	-	0.99836	0.99854

Данный способ позволяет увеличить точность для тестовой выборки, но при этом может снизить точность для тренировочной. В связи с этим данный метод повышения точности необходимо применять с аккуратностью. В дальнейшем при моделировании метод скользящего среднего не использовался.

Большую эффективность показало применение фильтра Савицкого-Голея к исходным данным. Результаты моделирования приведены в таблицах 3.5 и 3.6. В результате произошло значительное повышение точности: MSE снизилось с 0.0012 до 0.00026.

Таблица 3.5 - Метрики регрессионных моделей для тестовых выборок данных, к которым было применено сглаживание

Метрика	Linear Regression()	Lasso()	Ridge()	ElasticNetCV()	RidgeCV()	SGD Regressor()
MSE	0.00026	0.51752	0.00026	0.00044	0.00026	-
RMSE	0.01619	0.71939	0.01618	0.02085	0.01618	-
R2	0.99949	-	0.99949	0.99916	0.99949	-

Таким образом подтвердилось предположение, что сглаживание с использованием фильтра Савицкого-Голея способно снижать влияние шумов, искажающих точность измерения.

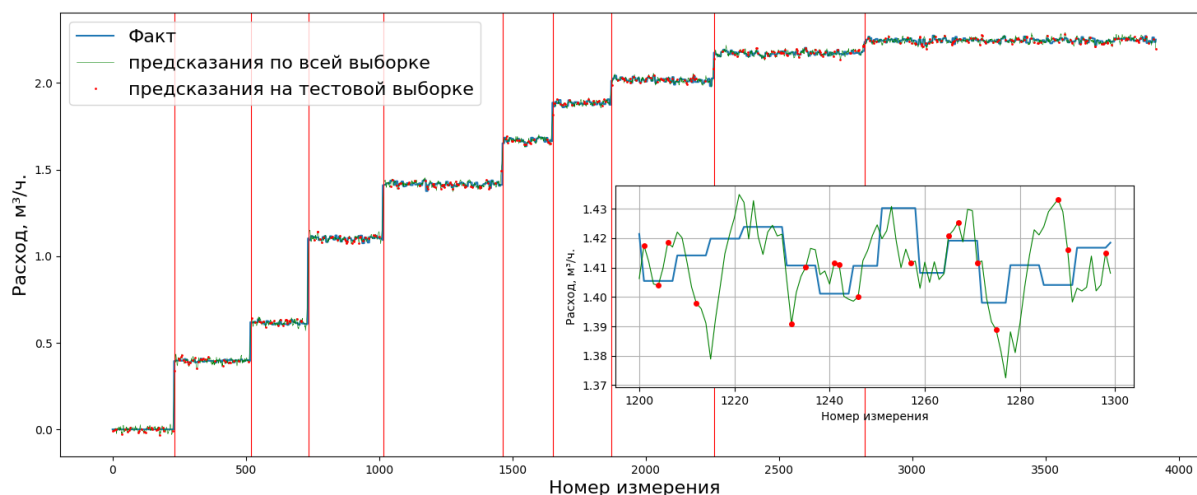


Рисунок 3.4 – Результаты предсказания модели LinearRegression для данных, к которым применено сглаживание Савицкого-Голея

Однако повторение со сглаженными данными моделирования, с выделением одного из режимов расхода в тестовую выборку, аналогичного повышения точности не произошло. MSE снизился всего на 0.0007 с 0.0026 до 0.0019 (см. таблицу 3.7).

Таблица 3.6 - Метрики модели LinearRegression для данных, к которым применено сглаживание Савицкого-Голея

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00026	0.00027	0.00027
RMSE	0.01619	0.01654	0.01651
R2	0.99949	0.99947	0.99947

Таблица 3.7 - Метрики модели LinearRegression для данных, к которым применено сглаживание Савицкого-Голея, где тестовой выборкой является один режим расхода

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00190	0.00022	0.00041
RMSE	0.04354	0.01470	0.02022
R2	-	0.99963	0.99921

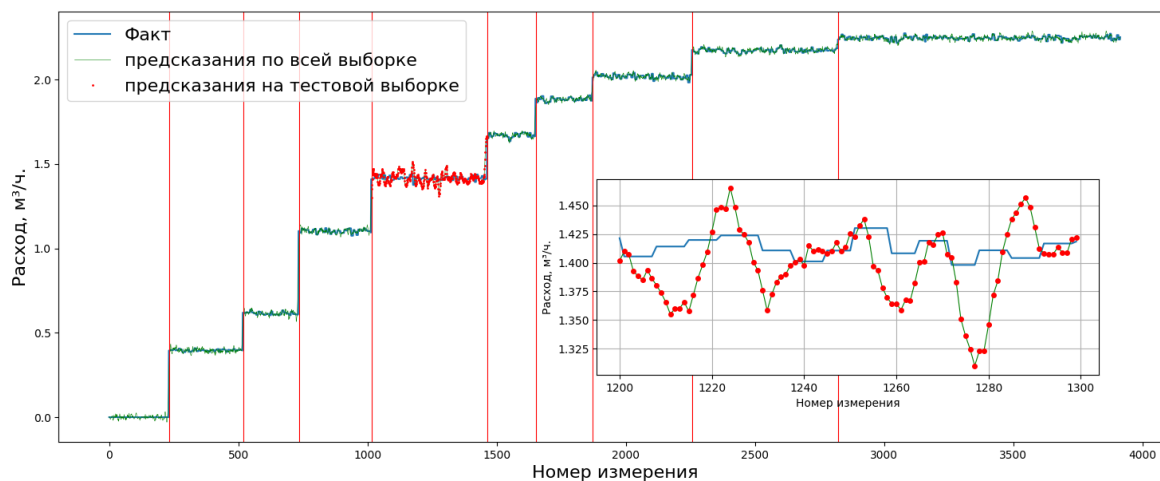


Рисунок 3.5 – Результаты предсказания модели LinearRegression для данных, к которым применено сглаживание Савицкого-Голея, где тестовой выборкой является один режим расхода

На основании результатов исследования линейных методов машинного обучения было принято решение о необходимости дальнейшего рассмотрения применимости других алгоритмов для цифровой обработки сигналов ультразвуковых расходомеров.

3.2.2 Результаты построения модели k-случайных соседей

При применении метода k-ближайших соседей для построение регрессионной модели для обработки сигналов ультразвуковых расходомеров получены неоднозначные результаты.

MSE для несглаженной выборки вдвое больше MSE линейной регрессии (см. таблицу 3.8). Кроме того, обращает на себя внимание рост ошибки для тестовой выборки с увеличением расхода жидкости. При этом размер выбросов сопоставим с шагом изменения расходов между режимами, на основании чего можно утверждать, что уменьшение шага изменения расходов при градуировке приведёт к росту ошибок такого характера.

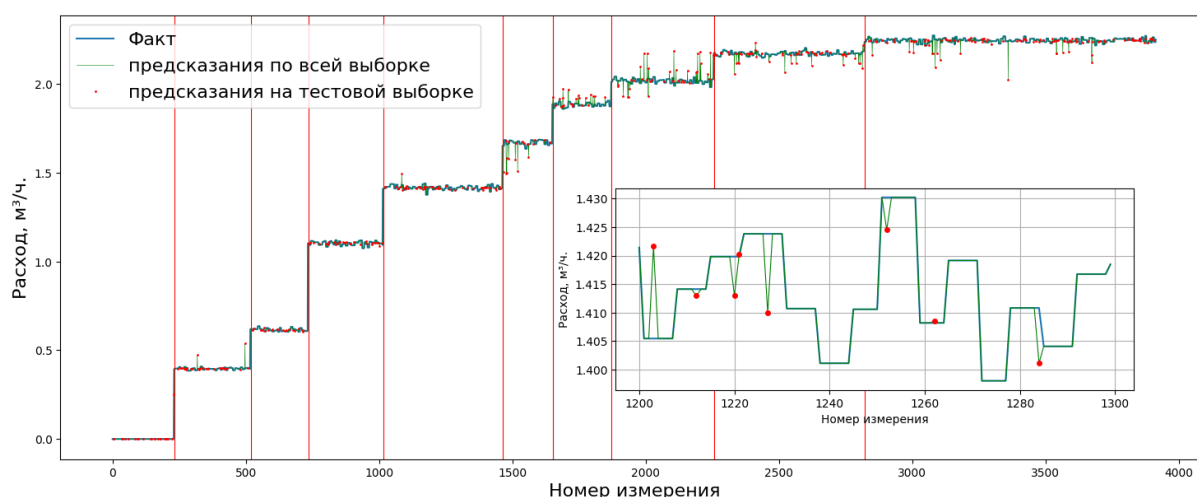


Рисунок 3.6 – Результаты предсказания модели KNeighborsRegressor для данных без применения сглаживания

Таблица 3.8 - Метрики модели KNeighborsRegressor для данных без применения сглаживания

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00217	0.00000	0.00021
RMSE	0.04655	0.00000	0.01463
R2	0.99582	1.00000	0.99959

При этом проведение моделирования, в котором тестовой выборкой выступает один режим с расходом 1,42 м³/ч. (с 1015 по 1464 измерение), выявило существенное ухудшение метрик (см. таблицу 3.9) из-а того, что

отсутствие расхода в выборке, на основании которой производится обучение модели, приводит к большим ошибкам в измерениях, так как модель присваивает значения расхода по ближайшим режимам расхода, учтённым в градуировке (рисунок 3.7).

Таблица 3.9 - Метрики модели KNeighborsRegressor для данных без применения сглаживания, где тестовой выборкой является один режим расхода

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.005144	0.00000	0.00590
RMSE	0.22681	0.00000	0.07681
R2	-	1.00000	0.98859

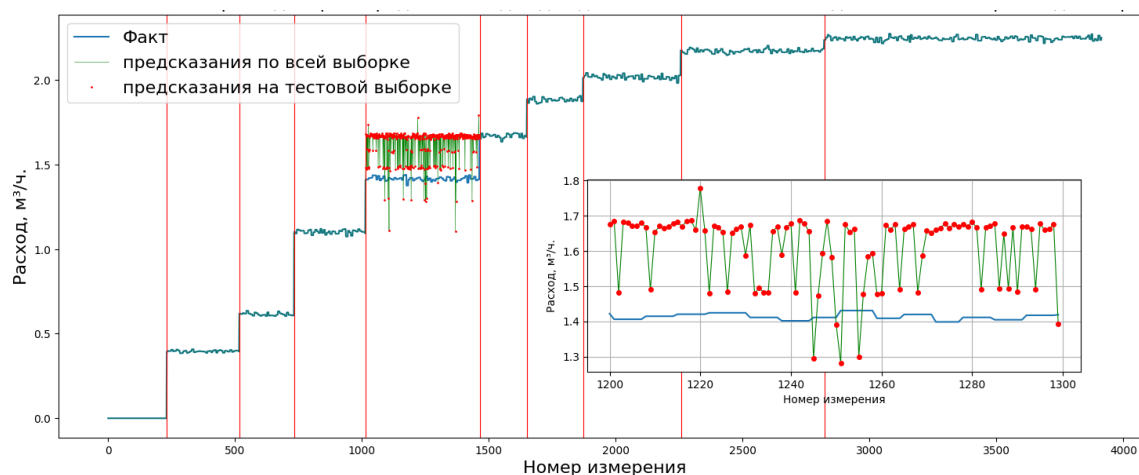


Рисунок 3.7 – Результаты предсказания модели KNeighborsRegressor для данных, где тестовой выборкой является один режим расхода

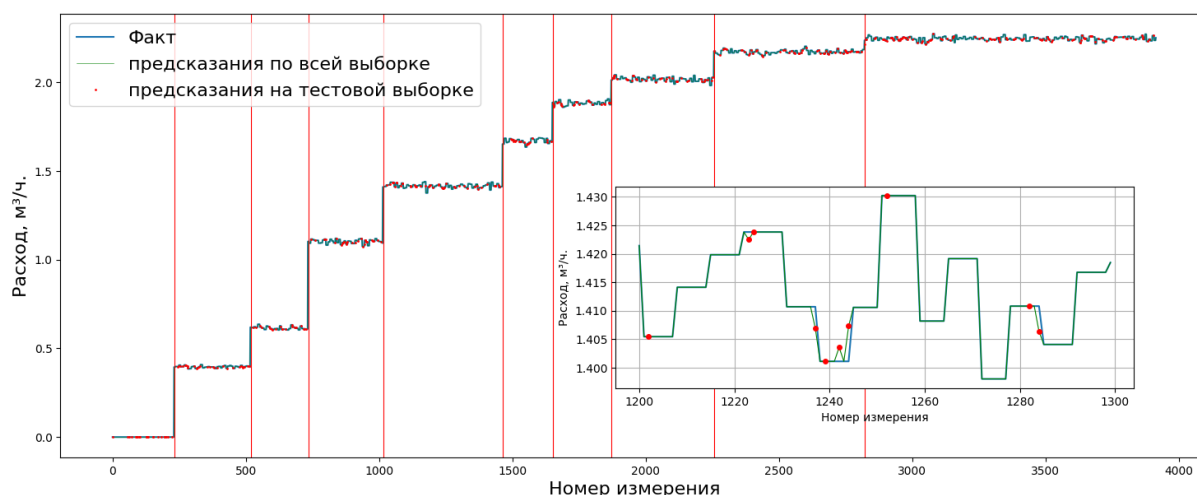


Рисунок 3.8 – Результаты предсказания модели KNeighborsRegressor для данных, к которым применено сглаживание Савицкого-Голея

Применение сглаживания с использованием Савицкого-Голея для конкретной исследуемой выборки решило проблему выбросов. Из таблицы 3.10 видно, что на тестовой выборке получен практически идеальный результат: MSE составила всего лишь 0.00003.

Таблица 3.10 - Метрики модели KNeighborsRegressor для данных, к которым применено сглаживание Савицкого-Голея

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.00003	0.00000	0.00000
RMSE	0.00544	0.00000	0.00171
R2	0.99994	1.00000	0.99999

Но проведение моделирования, в котором тестовой выборкой выступает один режим с расходом 1,42 м³/ч. (с 1015 по 1464 измерение), показывает, что применение сглаживания не исправляет ошибки, связанные с отсутствием расхода в обучающей выборке (рисунок 3.9).

Результаты моделирования представлены в таблице 3.11.

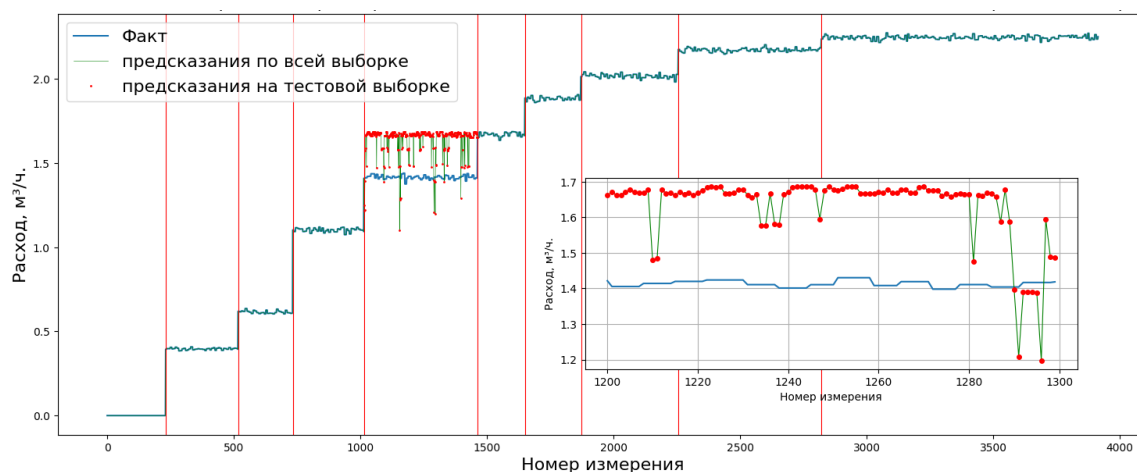


Рисунок 3.9 – Результаты предсказания модели KNeighborsRegressor для данных, к которым применено сглаживание Савицкого-Голея, где тестовой выборкой является один режим расхода

Таблица 3.11 - Метрики модели KNeighborsRegressor для данных, к которым применено сглаживание Савицкого-Голея, где тестовой выборкой является один режим расхода

Метрика	Тестовая выборка	Тренировочная выборка	Полная выборка
MSE	0.05626	0.00000	0.00645
RMSE	0.23719	0.00000	0.08032
R2	-	1.00000	0.98752

В итоге возникает неразрешимое противоречие:

- чем больше шаг изменения расхода, тем больше будет ошибок на режимах, не учтённых в процессе обучения;
- чем меньше шаг изменения расхода, тем больше будет ошибок на соседних режимах, учтённых в процессе обучения.

Учитывая все описанные выше проблемы, справедливым будет сделать вывод, что метод k-ближайших соседей не применим для целей обработки сигналов ультразвуковых расходомеров.

3.3 Исследование нейронных сетей

3.3.1 Результаты построения полносвязной нейронной сети

Рассмотрим сначала состоящую из трёх слоёв Dense полносвязную нейронную сеть, как самый простой и производительный вариант глубокого обучения (см. рисунок 3.10).

Так как использование во всех слоях функции активации gelu привело к застреванию в локальном оптимуме, в первом слое решено было заменить её на гиперболический тангенс tanh, что решило проблему.

Обучение длилось 512 эпох. За итоговую модель принималась модель, показавшая наилучшую функцию потерь на валидационной выборке.

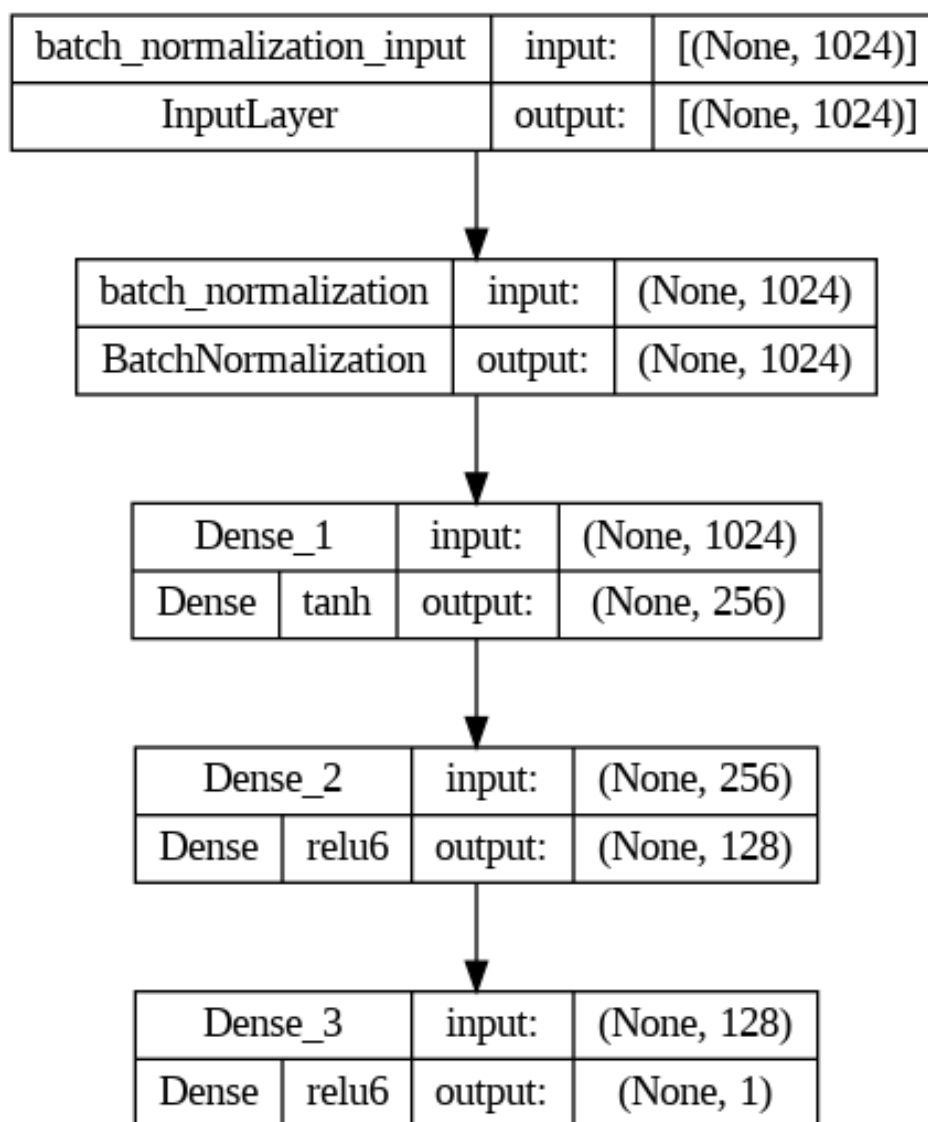


Рисунок 3.10 – Структура полносвязной нейронной сети

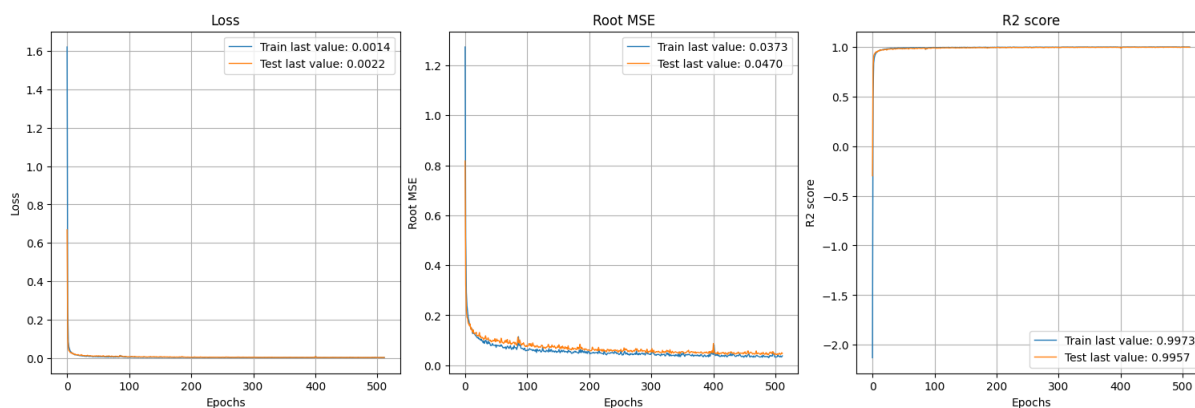


Рисунок 3.11 – Изменение функции потерь по эпохам обучения

Таблица 3.12 - Метрики модели полносвязной нейронной сети

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00045	0.0018	0.0022
RMSE	0.0214	0.0422	0.0466
R2	0.9991	0.9965	0.9958

Из таблицы 3.12 видно, что в результате построения модели полносвязной нейронной сети и подбора гиперпараметров не получилось выйти на приемлемые показатели точности расчёта. Кроме того, функция потерь для валидационной выборки на порядок выше тренировочной, что свидетельствует о наличии переобучения.

Также, при многократном запуске обучения модели имело место существенное различие в получаемых в результате метриках.

Также, как и для моделей линейной регрессии и k-ближайших соседей, было проведено моделирование, в котором тестовой выборкой выступил режим расхода 1,42 м³/ч.

Точность, как и в предыдущих случаях, для тестовой выборки существенно снизилась (см. таблицы 3.13 и 3.14). При этом ошибка есть и по отклонению среднего значения для режима расхода, и по росту дисперсии значений от среднего значения.

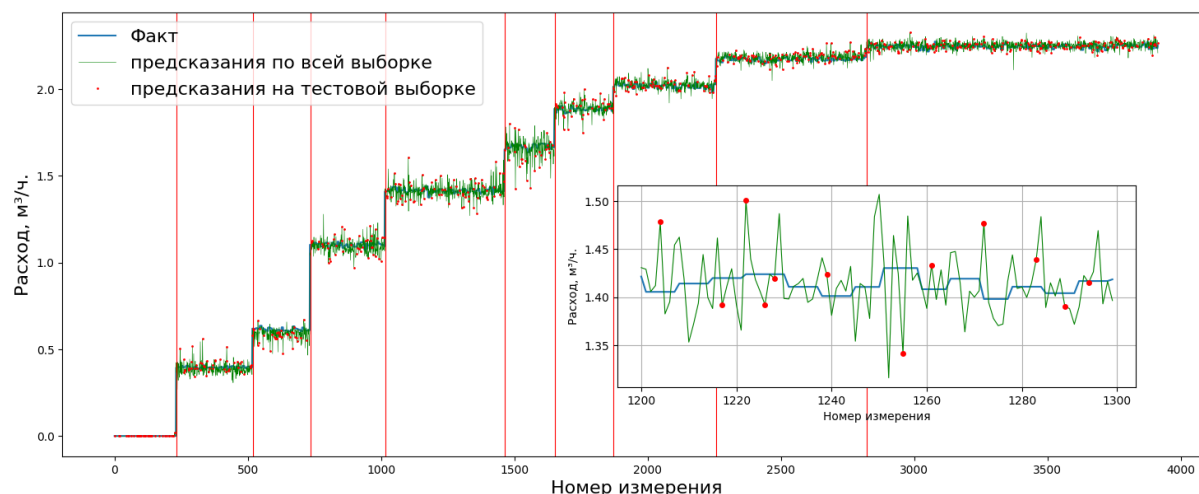


Рисунок 3.12 – Результаты предсказания модели полносвязной нейронной сети

Таблица 3.13 - Метрики модели полносвязной нейронной сети для данных, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00033	0.0012	0.013
RMSE	0.0181	0.0353	0.1140
R2	0.9994	0.9978	-

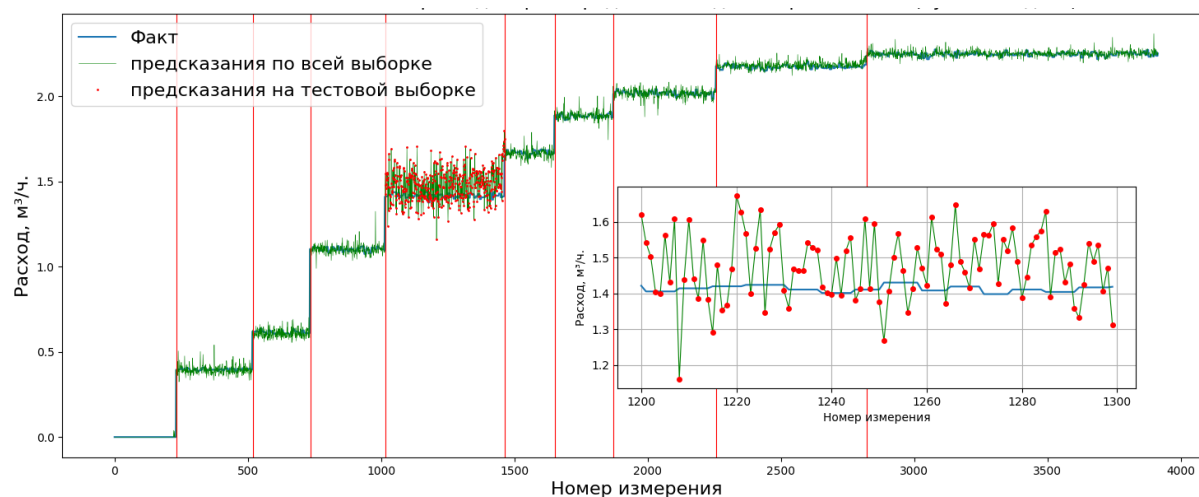


Рисунок 3.13 – Результаты предсказания модели полносвязной нейронной сети для данных, где тестовой выборкой является один режим расхода

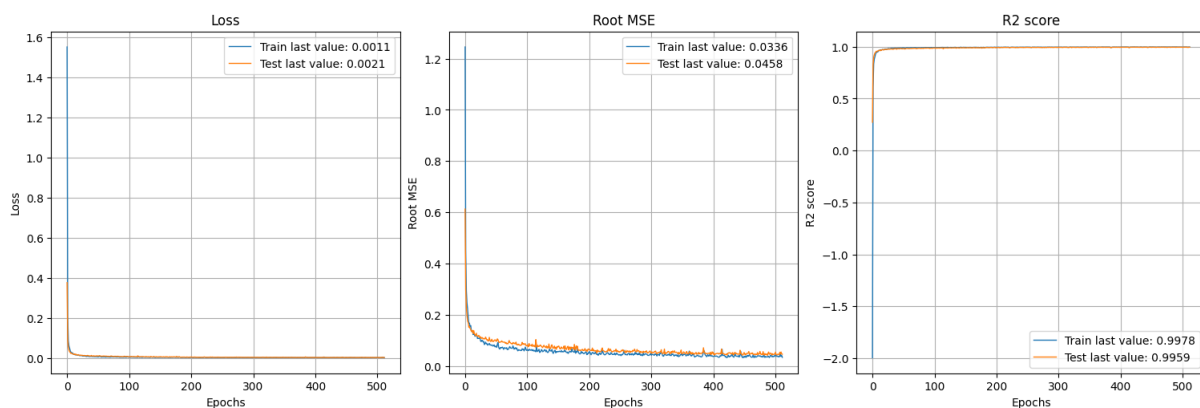


Рисунок 3.14 – Изменение функции потерь по эпохам обучения для данных, сглаженных фильтром Савицкого-Голея

Таблица 3.14. Метрики модели полносвязной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00048	0.0017	0.0021
RMSE	0.0218	0.0417	0.0457
R2	0.9991	0.9966	0.9966

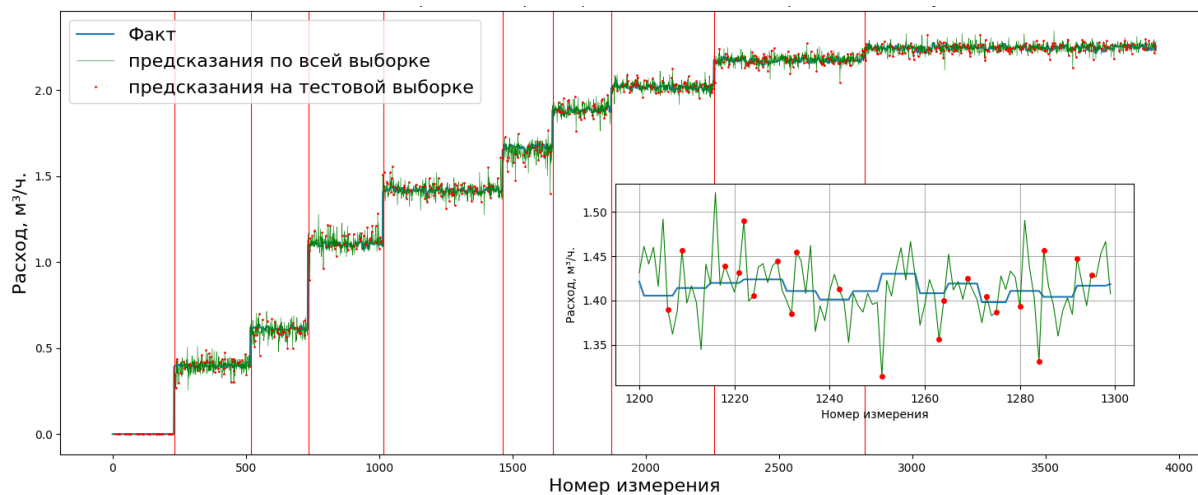


Рисунок 3.15 – Результаты предсказания модели полносвязной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

Таблица 3.15 - Метрики модели полносвязной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00038	0.0012	0.0159
RMSE	0.0195	0.0347	0.1261
R2	0.9993	0.9979	-

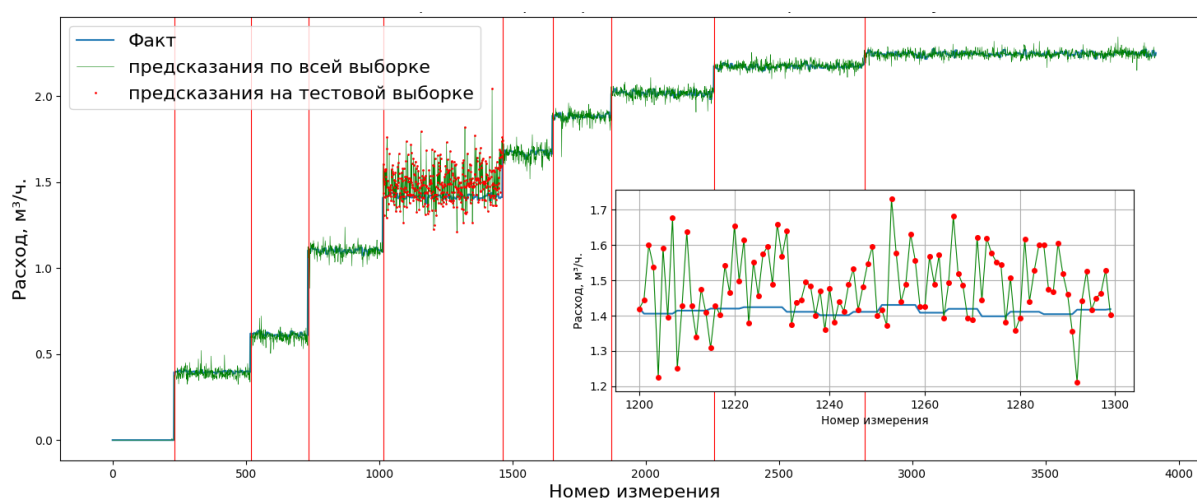


Рисунок 3.16 – Результаты предсказания модели полносвязной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

Применение сглаживания с помощью фильтра Савицкого-Голея не повлияло на улучшение точности модели, в том числе для данных, где один режим расхода выступал тестовой выборке и не участвовал в обучении модели.

Видно, что метрики, приведённые в таблицах 3.14 и 3.15, существенно не улучшились. То есть для полносвязной нейронной сети фактор сглаживания оказался несущественным. При этом существенная ошибка для модели, где тестовой выборкой выступает один режим расхода сохранилась.

3.3.2 Результаты построения свёрточной нейронной сети

Лучшие метрики показала модель свёрточной нейронной сети, состоящая из трёх стеков с одномерными слоями Conv1D с гиперболическим тангенсом (tanh) в качестве функции активации и Maxpooling1D. После стеков идёт слой регуляризации Dropout, слой преобразования в одномерный тензор Flatten и два полносвязных слоя Dense с функцией активации ReLU (рисунок 3.18).

Необходимо отметить, что регуляризация Dropout применялась в данной модели в незначительных объёмах, что свидетельствует об устойчивости модели к переобучению.

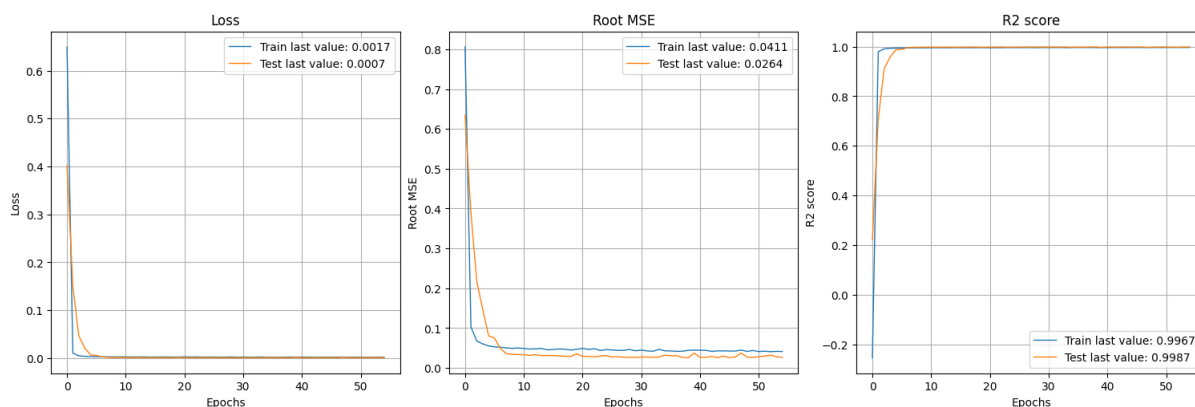


Рисунок 3.17 – Изменение функции потерь по эпохам обучения

Обучение длилось 64 эпохи. За итоговую модель принималась модель, показавшая наилучшую функцию потерь на валидационной выборке.

Таблица 3.16 - Метрики модели свёрточной нейронной сети

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00063	0.00064	0.00068
RMSE	0.0251	0.0254	0.0262
R2	0.9988	0.9987	0.9987

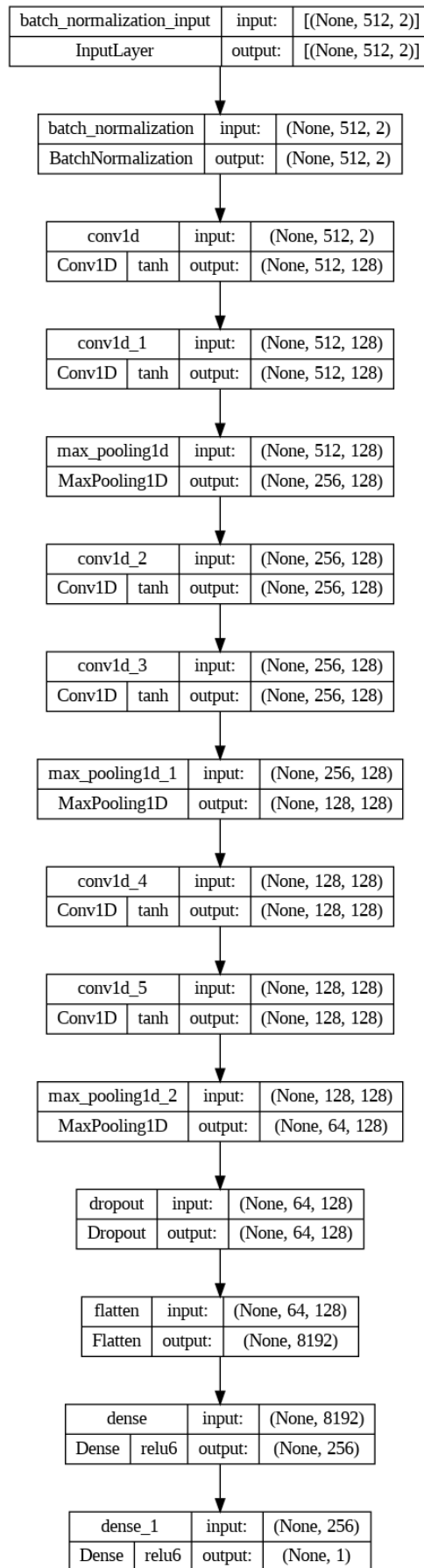


Рисунок 3.18 – Структура свёрточной нейронной сети

По результатам построения модели и анализа полученных метрик, приведённых в таблице 3.16, обращает на себя внимание, что метрики тестовой выборки сопоставимы с метриками тренировочной и валидационной, что также подтверждает отсутствие проблем с переобучением.

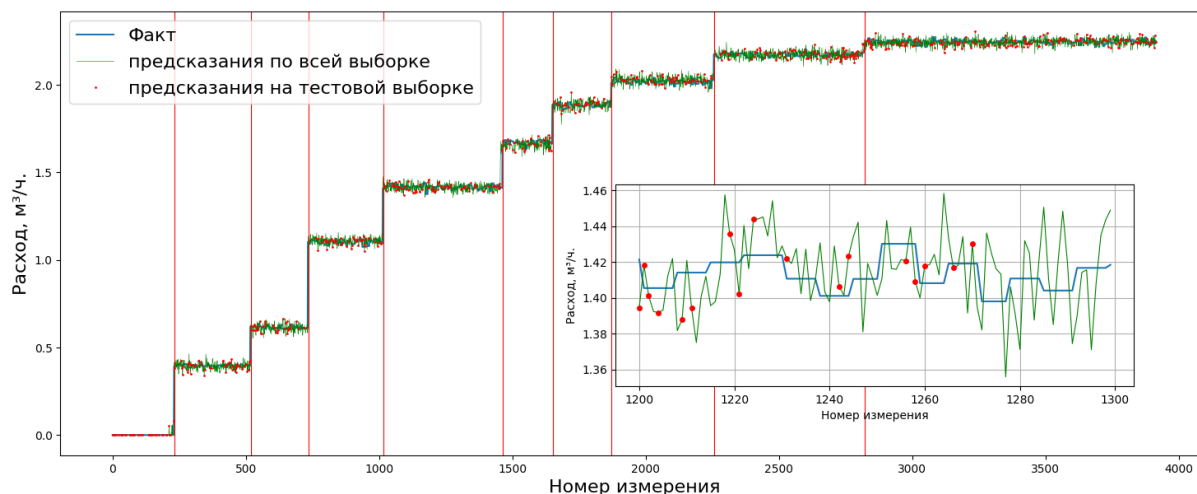


Рисунок 3.19 – Результаты предсказания модели свёрточной нейронной сети

Также, как и для других моделей, было проведено моделирование, в котором тестовой выборкой выступил режим расхода 1,42 м³/ч., не включённый в валидационную и тестовую выборки.

Таблица 3.17. Метрики модели свёрточной нейронной сети для данных, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00043	0.00052	0.0016
RMSE	0.0208	0.0229	0.0399
R2	0.9993	0.9991	-

Из таблицы 3.17 видно, что точность, как и в предыдущих случаях, для тестовой выборки существенно снизилась. Но для тестовых выборок, которые являются одним режимом расхода, не включённым в тренировочную и

валидационную выборки, показатель MSE 0.0016 является лучшим для всех изученных алгоритмов машинного обучения

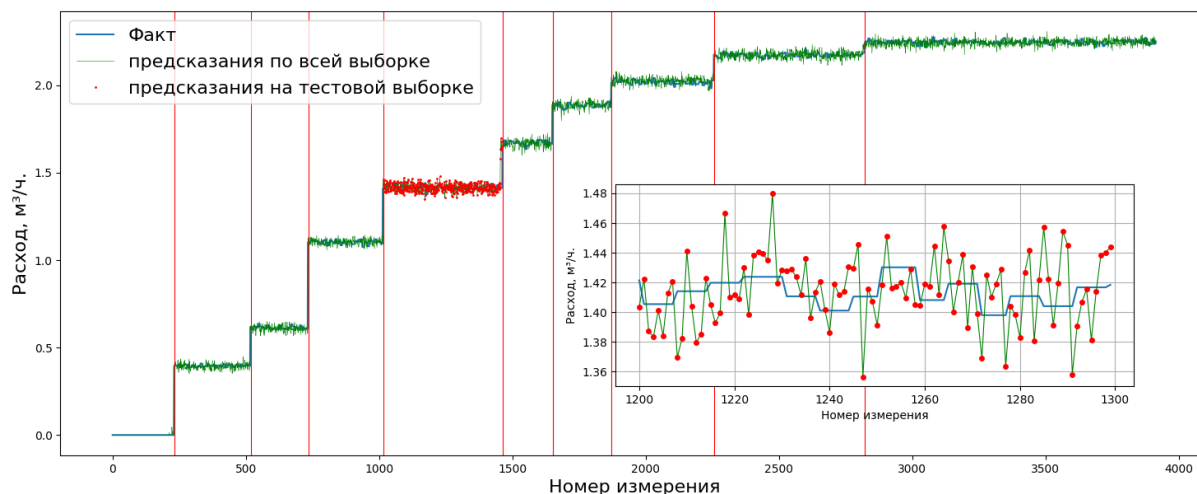


Рисунок 3.20 – Результаты предсказания модели свёрточной нейронной сети для данных, где тестовой выборкой является один режим расхода

Для сравнения с моделью линейной регрессии и анализа влияния шумов было проведено моделирование нейронной сети с данными, к которым предварительно было применено сглаживание с использованием фильтра Савицкого-Голея.

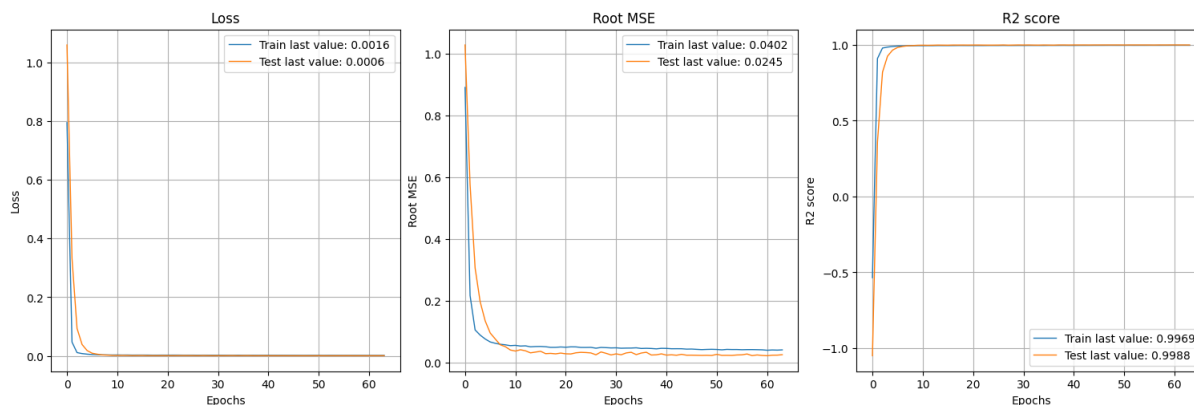


Рисунок 3.21 – Изменение функции потерь по эпохам обучения для данных, сглаженных фильтром Савицкого-Голея

Как следует из таблицы 3.18, прирост точности по метрикам после сглаживания не является таким же большим, как для линейной регрессии, но имеет место.

Таблица 3.18 - Метрики модели свёрточной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00038	0.00045	0.00036
RMSE	0.0196	0.0213	0.0191
R2	0.9993	0.9991	0.9993

По результатам построение модели обращает на себя внимание, что метрики тестовой выборки, приведённые в таблице 3.18, остаются не хуже метрик выборок тренировочной и валидационной, что также подтверждает отсутствие проблем с переобучением.

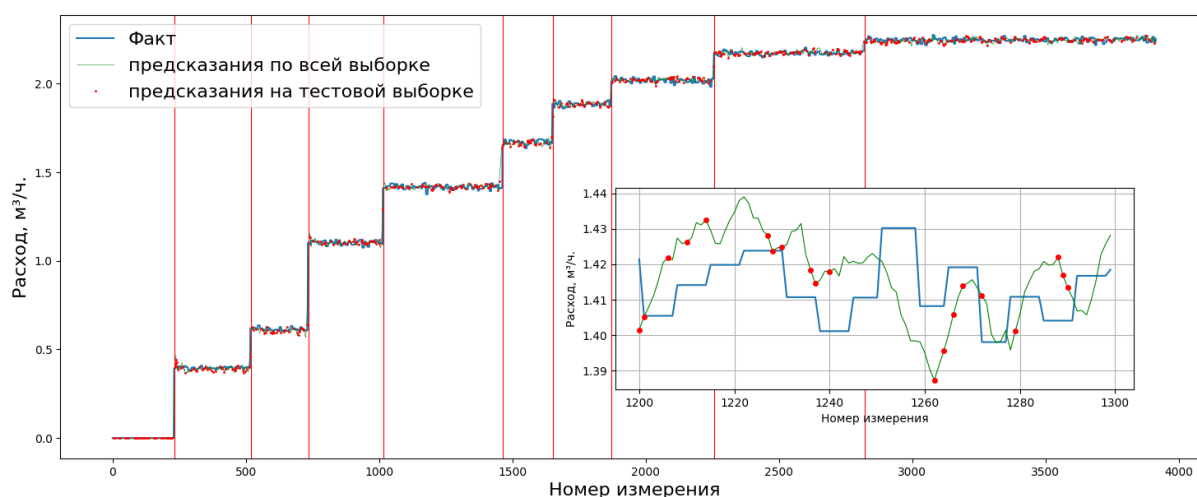


Рисунок 3.22 – Результаты предсказания модели свёрточной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

При моделировании, в котором тестовой выборкой выступил режим расхода 1,42 м³/ч., точность, как и в предыдущих случаях, для тестовой выборки существенно снизилась, о чём свидетельствуют полученные метрики, приведенные в таблице 3.19. Но при этом в случае свёрточной нейронной сети ошибка для режима расхода, отсутствующего в обучающей выборке, формируется в основном за счёт незначительного занижения среднего

значения по выборке, то для линейной регрессии ошибка вызвана увеличением дисперсии при сохранении среднего значения (см. рисунок 3.23).

Таблица 3.19 - Метрики модели свёрточной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00030	0.00032	0.0016
RMSE	0.0173	0.0178	0.0402
R2	0.9995	0.9994	-

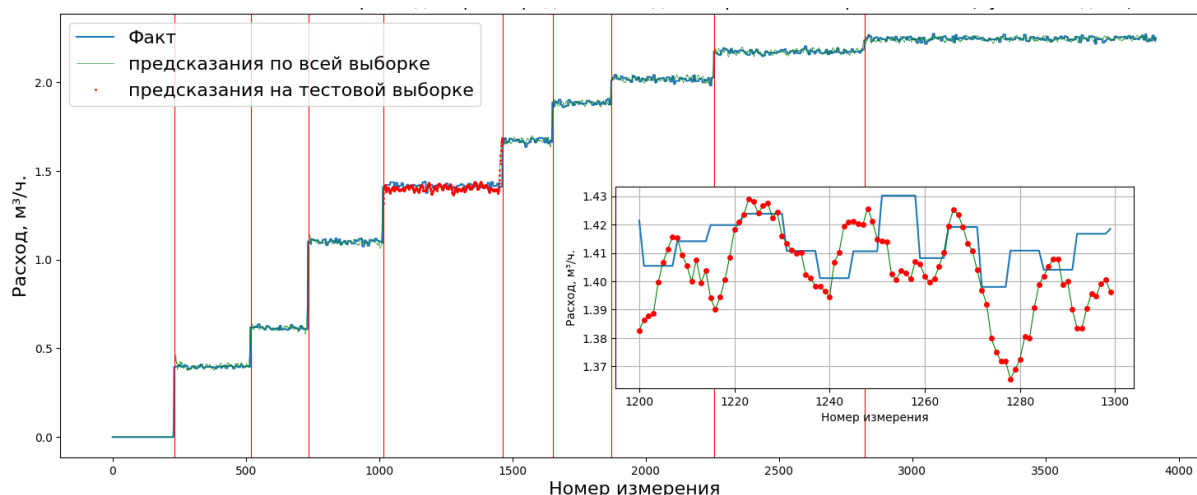


Рисунок 3.23 – Результаты предсказания модели свёрточной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

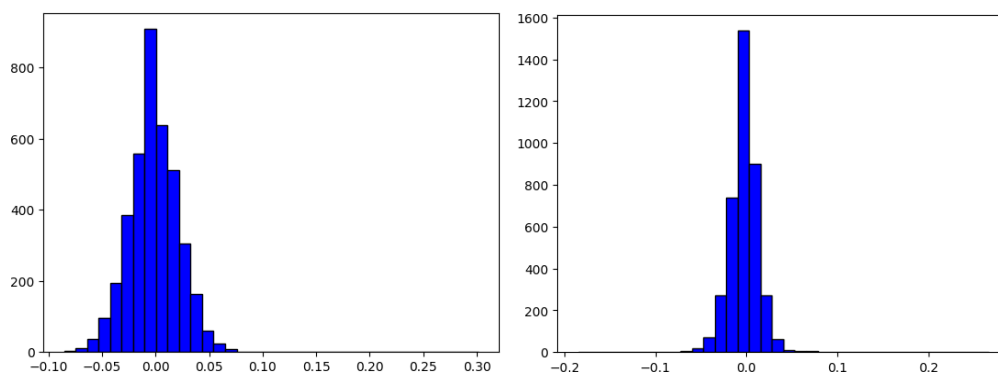


Рисунок 3.24 – Гистограммы ошибок предсказания для моделей без сглаживания (слева) и со сглаживанием (справа)

3.3.3 Результаты построения рекуррентной нейронной сети

Для рекуррентных нейронных сетей использован тип GRU. Лучшие метрики показала рекуррентная нейронная сеть GRU, состоящая из трёх слоёв с применением функции активации `relu6` к первым двум и `tanh` – к последнему.

Обучение длилось 64 эпохи. За итоговую модель принималась модель, показавшая наилучшую функцию потерь на валидационной выборке.

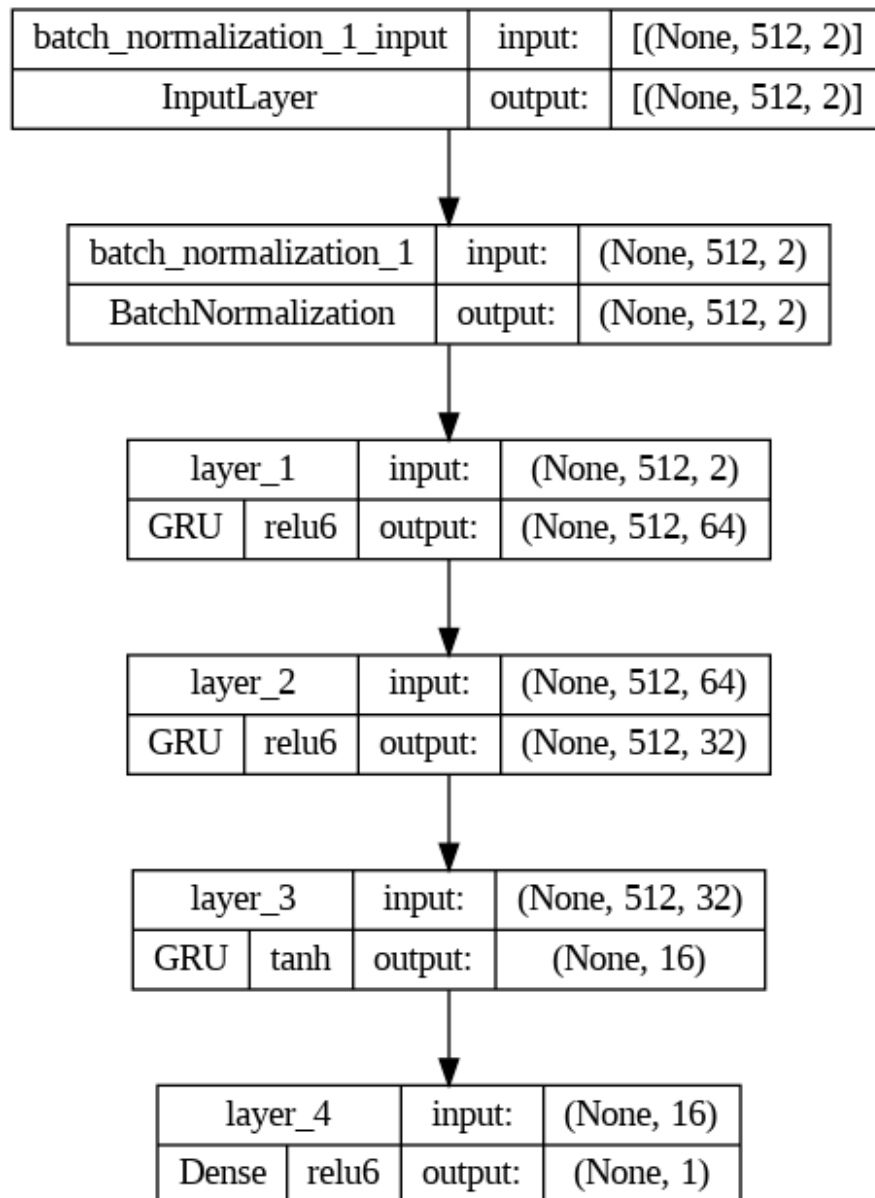


Рисунок 3.25 – Структура рекуррентной нейронной сети

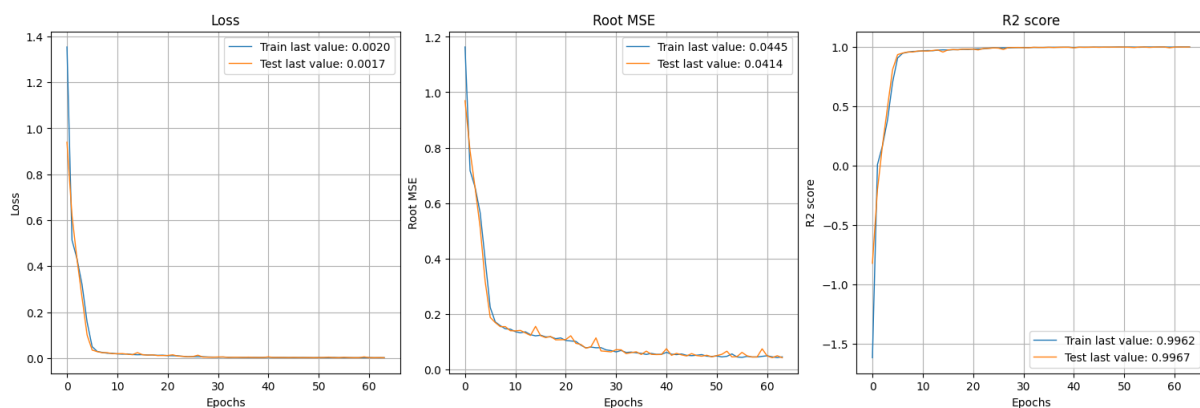


Рисунок 3.26 – Изменение функции потерь по эпохам обучения

Таблица 3.20 - Метрики модели рекуррентной нейронной сети

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.0015	0.0017	0.0017
RMSE	0.0389	0.0414	0.0416
R2	0.9971	0.9967	0.9967

Из полученных метрик, приведённых в таблице 3.20, следует, что результат модели существенно хуже, чем у свёрточной нейронной сети.

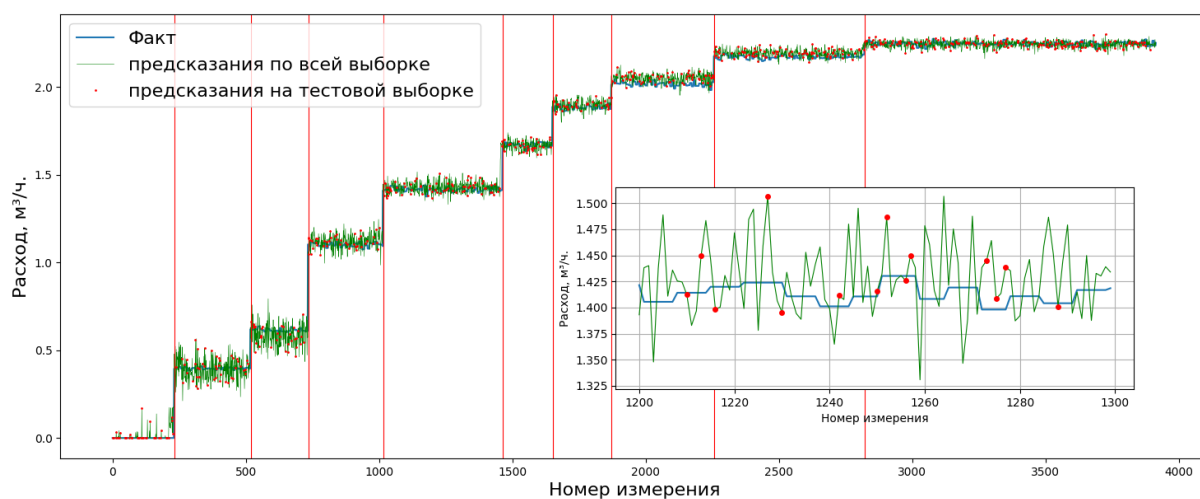


Рисунок 3.27 – Результаты предсказания модели рекуррентной нейронной сети

Таблица 3.21 - Метрики модели свёрточной нейронной сети для данных, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.0014	0.0015	0.0065
RMSE	0.0374	0.0388	0.0805
R2	0.9976	0.9974	-

Обращает внимание, что точность построенной модели рекуррентной нейронной сети растёт с увеличением расхода, что видно на рисунке 3.27.

При моделировании рекуррентной сети для данных, где тестовой выборкой является один режим расхода, рост погрешности в тестовой выборке усугубляется тем, что модель изначально менее точная, чем ранее рассмотренные.

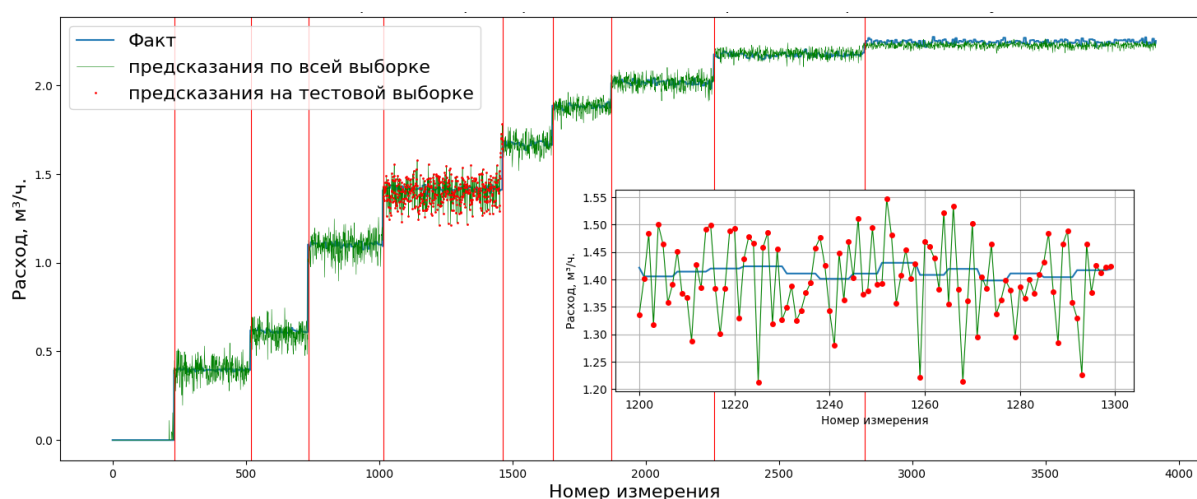


Рисунок 3.28 – Результаты предсказания модели рекуррентной нейронной сети для данных, где тестовой выборкой является один режим расхода

Применение сглаживания с использованием фильтра Савицкого-Голея даёт существенный прирост точности и убирает зависимость погрешности от расхода.

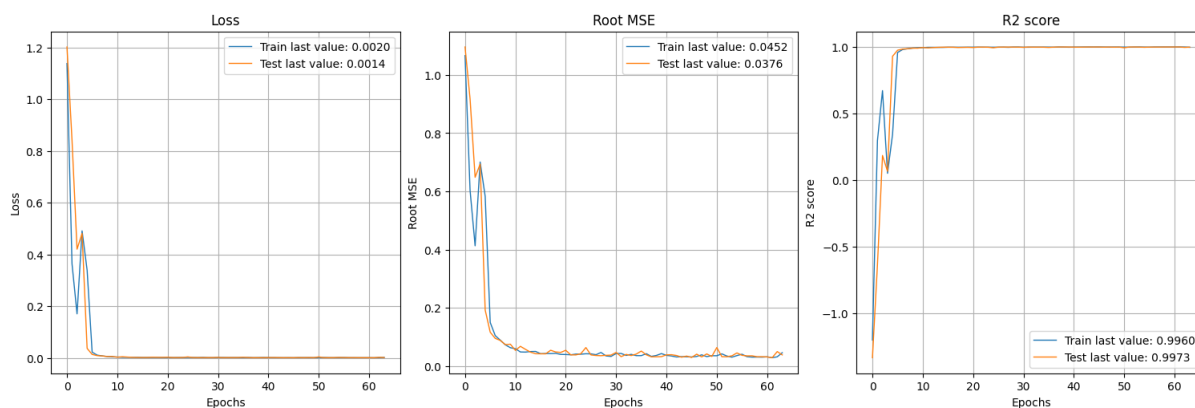


Рисунок 3.29 – Изменение функции потерь по эпохам обучения для данных, сглаженных фильтром Савицкого-Голея

Таблица 3.22 - Метрики модели рекуррентной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00067	0.00079	0.00067
RMSE	0.0260	0.0282	0.0258
R2	0.9987	0.9985	0.9987

Из таблиц 3.20 и 3.22 видно, что как в случае свёрточной нейронной сети, метрики тестовой выборки для рекуррентной не хуже метрик тренировочной и валидационной. Но всё же на графике видно, что процесс обучения был нестабильным и по валидационной выборке наблюдаются всплески переобучения.

При моделировании, в котором тестовой выборкой выступил режим расхода $1,42 \text{ м}^3/\text{ч.}$, имеет место уже зафиксированный ранее эффект роста ошибки. Метрики для модели приведены в таблице 3.23.

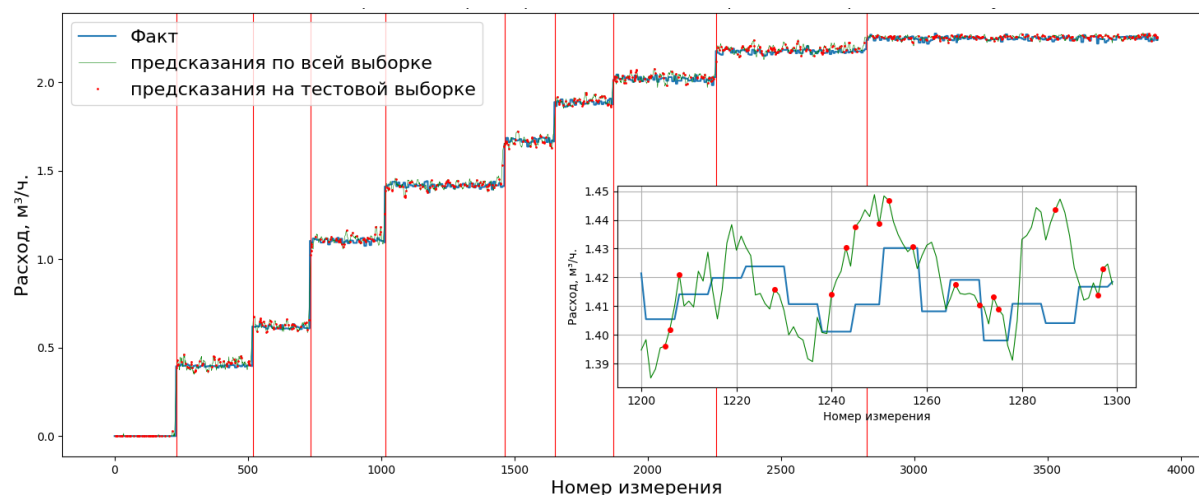


Рисунок 3.30 – Результаты предсказания модели рекуррентной нейронной сети для данных, сглаженных фильтром Савицкого-Голея

Таблица 3.23. Метрики модели свёрточной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

Метрика	Тренировочная выборка	Валидационная выборка	Тестовая выборка
MSE	0.00052	0.00045	0.0026
RMSE	0.0228	0.0214	0.0509
R2	0.9991	0.9992	-

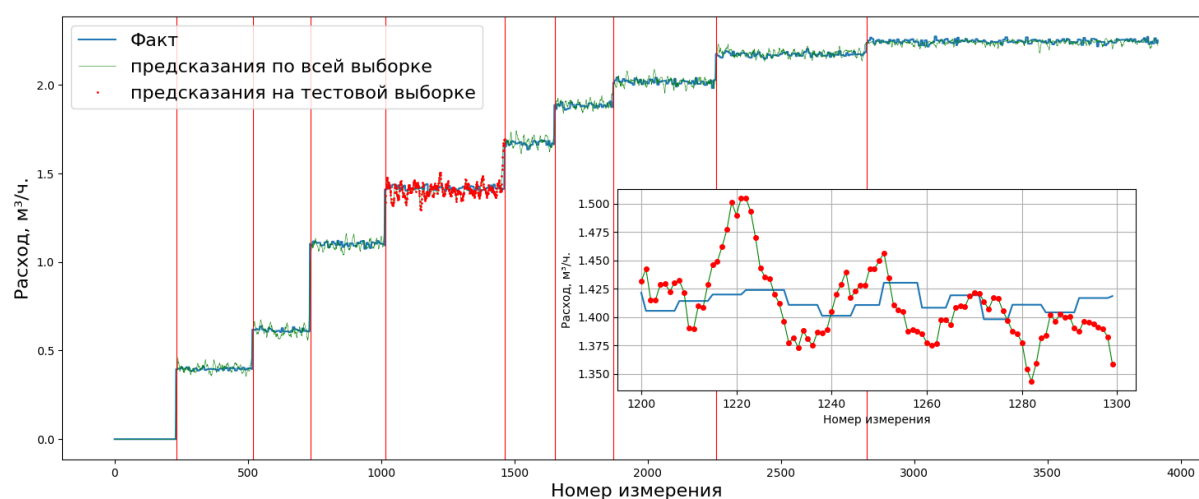


Рисунок 3.31 – Результаты предсказания модели рекуррентной нейронной сети для данных, сглаженных фильтром Савицкого-Голея, где тестовой выборкой является один режим расхода

3.4 Выводы из главы 3

В рамках главы 3 разработаны модели машинного обучения, определённые как подходящие для целей исследования в главе 2.

При применении алгоритмов машинного обучения погрешность измерения не зависит от величины расхода, что решает проблему, связанную с калибровкой и градуировкой ультразвуковых расходомеров и применением корректирующего гидравлического коэффициента.

При отсутствии режима расхода в обучающей выборке возникает рост погрешности за счёт роста дисперсии без изменения средней величины расхода.

Из рассмотренных алгоритмов наилучшие метрики у линейной регрессии: MSE 0.00026, RMSE 0.01618, R2 0.9995, и у свёрточной нейронной сети: MSE 0.00036, RMSE 0.0191, R2 0.9993.

ЗАКЛЮЧЕНИЕ

В рамках проведённой работы выполнено исследование методов обработки сигналов ультразвуковых расходомеров с использованием алгоритмов машинного обучения, направленных на повышение точности измерения за счёт исключения влияния на расчёт расхода гидравлического корректирующего коэффициента.

В результате оптимальные метрики показали линейная регрессия: MSE 0.00026, RMSE 0.01618, R^2 0.9995), что соответствует погрешности измерения в 0,9%, и свёрточная нейронная сеть: MSE 0.00036, RMSE 0.0191, R^2 0.9993, что соответствует погрешности измерения в 1,1%,

Применение методов машинного обучения решает проблему увеличения погрешности измерения при отклонении значения расхода от величины, при которой осуществлялась калибровка, и значительно упрощает процесс градуировки оборудования.

Рост погрешности для режимов расхода, не участвовавших в обучающей выборке, свидетельствует о том, что при градуировке ультразвуковых расходомеров с использованием алгоритмов машинного обучения необходимо обеспечивать все необходимые режимы расхода при сборе тренировочных данных.

По результатам работы можно сделать вывод, что алгоритмы машинного обучения имеют большие перспективы использования в ультразвуковой расходомерии. Этому способствуют развитие отрасли информационных технологий и увеличение вычислительных мощностей оборудования.

Для дальнейшего развития направления машинного обучения в качестве алгоритма обработки сигналов датчиков ультразвуковых расходомеров необходимо провести дополнительные исследования на данных аналогичных экспериментов с трубопроводами других диаметров и изготовленных из других материалов.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Маркушин, М. Некоторые проблемы современной метрологии в микроэлектронике / М. Маркушин, В. Мартынов // Электроника: Наука. Технология. Бизнес., 2020, №2 (193), с. 80-91.
2. Кастелан, Б. Более быстрые измерения с использованием ИИ / Б. Кастелан, Р. Альфельд // Control Engineering Russia. – 2023. – № 3. – С.64-65.
3. Кремлевский, П.П. Расходомеры и счетчики количества вещества. Книга 2 / П.П. Кремлевский –М.: Политехника, 2004. – 411 с.
4. Малыгин, И.В. Применение методов машинного обучения для классификации радиосигналов / И. В. Малыгин, С.А. Бельков, А.Д. Тарасов, М.Р. Усвятцов // Труды МАИ. 2017. № 96. [Электронный ресурс]. URL: <http://trudymai.ru/published.php?ID=85797> (дата обращения 15.04.2024).
5. Бахтин, А.А. Система распознавания модуляций сигналов на основе нейронной сети с использованием ПЛИС / А.А. Бахтин, А.С. Волков, А.В. Солодков, И.А. Свиридов // Труды МАИ. – 2021. - №121. – С. 1-25.
6. Хлыбов, А.А. Оценка поврежденности металла с использованием неразрушающего контроля и подходов искусственного интеллекта / А.А. Хлыбов, Ю.Г. Кабалдин, М.С. Аносов, Д.А. Шатагин, Д.А. Рябов // Тенденции развития науки и образования, 2021, №76-1, С. 122-125.
7. Яблоков, А.Е. Диагностика технологического оборудования по интегральным характеристикам вибрации и звука с использованием методов машинного обучения / А.Е. Яблоков, Б.Н. Федоренко, И.Г. Благовещенский, Е.А. Олышанова // Проблемы современной науки и образования. 2019. № 7 (140). С. 25-29.
8. Власов, А.И. Применение систем искусственного интеллекта в электрокардиографии / А.И. Власов, С.Э. Гюльмалиева, Ю.С. Либбер, С. Абдулкадер // Нейрокомпьютеры: разработка, применение. 2022. Т. 24. № 4. С. 36-52.

9. Бердибаева, Г.К. Классификация звуков астматического дыхания с использованием нейронных сетей / Г.К. Бердибаева, О.Н. Бодин, Д.С. Фирсов // Измерение. Мониторинг. Управление. Контроль. - 2018. - №2(24). - С.86-90.
10. Кузин, А.Ю. Искусственный интеллект как один из элементов цифровой трансформации в метрологии / А.Ю. Кузин, А.Н. Крошкин // Инновационное приборостроение, 2022, №1, том 1, С. 27-35.
11. Wan, Thiam. Machine Learning Applied to Ultrasonic Flow Meters for measuring Dilute, Turbulent Water-Bentonite Suspension Flow / Thiam Wan, Hon Chung Lau, Wai Lam Loh // ResearchGate. – 2020. – 9. – p. 1-17.
12. Santhosh, K.V. An Intelligent Flow Measurement Technique using Ultrasonic Flow Meter with Optimized Neural Network / K.V. Santhosh, B.K. Roy // International Journal of Control and Automation. – 2002. – №4. – vol.5. – p. 185-195.
13. Yazdanshenasshad, B. Neural-network-based error reduction in calibrating utility ultrasonic flow meters / B. Yazdanshenasshad, M.S. Safizadeh // Flow Measurement and Instrumentation. – 2018. – vol.64. – p. 54-63.
14. Руководство по эксплуатации ЭТЛН.421364.003РЭ. Расходомеры-счётчики жидкости ультразвуковые многолучевые. Эталон РМ. – ООО "НПП «Эталон Энерго», 2012. – 33 с.
15. Шарапов, В.М. Электроакустические преобразователи / В.М. Шарапов, И.Г. Минаев, Ж.В. Сотула, Л.Г. Куницкая. - Москва: Техносфера, 2013. – 296 с.
16. Калмыков, А.А. Сравнение алгоритмов измерения задержки сигналов ультразвуковых расходомеров / А.А. Калмыков, М.В. Ронкин // Материалы 26-й Междунар. Крымск. конф. «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2016). Москва; Минск; Севастополь. 2016. Т. 10. С. 2330-2336.
17. Ронкин, М.В. Исследование особенностей использования частотно модулированных сигналов в ультразвуковых расходомерах / М.В. Ронкин // Материалы 26-й Междунар. Крымск. конф. «СВЧ-техника и

телекоммуникационные технологии» (КрыМиКо'2016). Москва; Минск; Севастополь. 2016. Т. 10. С. 2323-2329.

18. Соломичев, Р.И. Методы повышения метрологической надежности ультразвуковых расходомеров в условиях воздействия дестабилизирующих факторов / Р.И. Соломичев, А.Н. Слонько, В.А. Дубинин // Газовая промышленность, 2019, № S2 (786), С. 112-117.

19. Wang, Yi Fan. Applications of machine learning and data mining methods for advanced equipment control and process control / Yi Fan Wang, Yuri Ishizaki, Kazuho Sakurai, Luyi Li, Takuya Nakagawa, Sumika Arima // Conference: AEC/APC Symposium Asia 2015, p.1-4.

20. Краснов, П.В. Реализация метода «скользящее среднее» в системе автоматического управления / П.В. Краснов, С.А. Никульшин, Е.Ю. Решетова // Шаг в науку, 2017, №4, с. 48-53.

21. Ляхов, А.А. Применение фильтров Савицкого-Голея для обработки вольтамперных характеристик зондов Ленгмюра / А.А. Ляхов, В.В. Шкуркин // Вестник Омского университета 2012. Вып. 4. С. 72-76.

22. Каламбет, Ю. А. Оптимизация параметров линейного сглаживания хроматографических пиков / Ю.А. Каламбет // Научное приборостроение. 2019. Т. 29, № 3. С. 51-60.

23. Чжэн, Элис. Машинное обучение. Конструирование признаков: принципы и техники для аналитиков / Элис Чжен, Аманда Казари. – Москва, Эксмо, 2022. – 240 с.

24. Масис, С. Интерпретируемое машинное обучение на Python. / С. Масис. - СПб.: БХВ-Петербург, 2023. – 640 с.

25. Barandas, M. TSFEL: Time Series Feature Extraction Library / Marília Barandas, Duarte Folgado, Letícia Fernandes, Sara Santos, Mariana Abreu, Patrícia Bota, Hui Liu, Tanja Schultz, Hugo Gamboa // SoftwareX. – 2020. - №11. – 100456.

26. Мелкумова, Л.Э. Сравнение методов Ридж-регрессии и LASSO в задачах обработки данных / Л.Э. Мелкумова, С.Я. Шатских // Сборник трудов III международной конференции и молодежной школы «Информационные

технологии и нанотехнологии» (ИТНТ-2017) - Самара: Новая техника, 2017. - С.1748-1755.

27. Шолле, Ф. Глубокое обучение на Python / Ф. Шолле. – СПб.: Питер, 2023. – 576 с.

28. Траск, Эндрю. Грокаем глубокое обучение / Эндрю Траск – СПб.: Питер, 2024. – 352 с.

29. Ronkin, M. Novel FMCW-interferometry method testing on an ultrasonic clamp-on flowmeter / M. Ronkin, A. Kalmykov, K. Zeyd // IEEE Sensors Journal. – 2020. – Т. 20. – №11. – p. 6029-6037.

30. Пустовалов, Г.Е. Погрешности измерений / Г.Е. Пустовалов. – М.: МГУ им. М.В. Ломоносова, 2001. – 17 с.