

УТВЕРЖДАЮ
Декан факультета

_____ Страхов С.Ю.

«___» _____ 20__

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ

Направление/специальность подготовки	24.05.06 Системы управления летательными аппаратами
Специализация/профиль/программа подготовки	Системы управления ракет
Уровень высшего образования	Специалитет
Форма обучения	Очная
Факультет	И Информационные и управляющие системы
Выпускающая кафедра	ИЗ Системы управления и компьютерные технологии
Кафедра-разработчик рабочей программы	ИЗ Системы управления и компьютерные технологии

КУРС	СЕМЕСТР	ОБЩАЯ ТРУДОЁМКОСТЬ (ЗАЧЕТНЫХ ЕДИНИЦ)	ЧАСЫ (по наличию видов занятий)									ВИД ПРОМЕЖУТОЧНОГО КОНТРОЛЯ
			ОБЩАЯ ТРУДОЁМКОСТЬ	АУДИТОРНЫЕ ЗАНЯТИЯ				САМОСТОЯТЕЛЬНАЯ РАБОТА				
				ВСЕГО	ЛЕКЦИИ	ЛАБОРАТОРНЫЙ ПРАКТИКУМ	ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	ВСЕГО	КУРСОВОЙ ПРОЕКТ	КУРСОВАЯ РАБОТА	ДРУГИЕ ВИДЫ САМОСТ. РАБОТЫ	
4	7	3	108	34	0	0	34	74	0	0	74	зач.

ЛИСТ СОГЛАСОВАНИЯ

РАБОЧАЯ ПРОГРАММА СОСТАВЛЕНА В СООТВЕТСТВИИ С ТРЕБОВАНИЯМИ ФЕДЕРАЛЬНОГО
ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА ВЫСШЕГО ОБРАЗОВАНИЯ (ФГОС ВО)

24.05.06 Системы управления летательными аппаратами

год набора группы: 2026

Программу составил:

Кафедра ИЗ Системы управления и компьютерные технологии
Кострыгин Даниил Геннадьевич, преподаватель

Программа рассмотрена
на заседании кафедры-разработчика
рабочей программы **ИЗ Системы управления и компьютерные технологии**

Заведующий кафедрой Сырцев А.Н., д.воен.н., снс

Программа рассмотрена
на заседании выпускающей кафедры

ИЗ Системы управления и компьютерные технологии

Заведующий кафедрой Сырцев А.Н., д.воен.н., снс

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ

Разделы рабочей программы

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО
3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ
4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Приложения к рабочей программе дисциплины

- Приложение 1. Аннотация рабочей программы
- Приложение 2. Технологии и формы обучения
- Приложение 3. Фонды оценочных средств

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является формирование следующих компетенций:

ПК-И1 — владеет технологиями и инструментами искусственного интеллекта для решения задач профессиональной деятельности

ПК-И2 — способен применять цифровые производственные системы в области профессиональной деятельности

Формированию компетенций служит достижение следующих результатов образования:

ПК-И1

знания:

Принципы обработки больших данных (Big Data) в реальном времени;

Архитектуру распределённых вычислительных систем (Hadoop, Spark, Kafka);

Методы машинного обучения для анализа телеметрии и диагностики сложных технических систем;

Инструменты для предиктивной аналитики и оптимизации управления;

Особенности работы с потоковыми данными в авиационных системах;

Методики очистки, нормализации и верификации данных в условиях зашумлённости и неполноты;

умения:

Применять распределённые вычисления для анализа больших объёмов исторических и потоковых данных;

Выбирать и настраивать модели машинного обучения для задач прогнозирования отказов и оптимизации параметров;

Интерпретировать результаты анализа в терминах предметной области (надёжность, точность, ресурс);

Применять классические методы математической статистики для анализа больших массивов данных (описательная статистика, проверка гипотез, регрессионный анализ);

навыки:

Практической работы с PySpark, Kafka, MLlib для обработки телеметрии ЛА;

Реализации пайплайнов предиктивного обслуживания на основе реальных данных испытательных стендов;

Визуализации результатов анализа для поддержки принятия инженерных решений;

Использования библиотек статистического анализа (SciPy, StatsModels) в распределённой среде.

ПК-И2

знания:

Архитектуру цифровых производственных систем и цифровых двойников на всех этапах жизненного цикла продукции;

Стандарты и протоколы интеграции данных в авиационной отрасли;

Методы статистического управления качеством (SPC) и анализа процессов в цифровых системах;

Принципы построения систем поддержки принятия решений на основе больших данных;

Требования к надёжности, безопасности и аудиту данных в критических системах управления;

умения:

Проектировать компоненты цифровых двойников для анализа телеметрии и моделирования поведения систем управления;

Настраивать интеграцию данных от испытательного оборудования в единое информационное пространство;

Применять аналитические модели для оптимизации производственных процессов и повышения точности изготовления узлов;

навыки:

Работы с IoT-платформами для приёма и визуализации данных с датчиков;

Внедрения решений Big Data в контуры управления качеством и производственным планированием;

Разработки прототипов систем предиктивного обслуживания на основе потоковой аналитики;

Построения контрольных карт и анализа воспроизводимости процессов (R&R).

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ** является дисциплиной **части, формируемой участниками образовательных отношений блока 1**, программы подготовки по направлению *24.05.06 Системы управления летательными аппаратами*.

Содержание дисциплины является логическим продолжением дисциплин: **ТЕХНОЛОГИИ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ПРОГРАММНЫХ СИСТЕМ, МЕТОДЫ ПРИНЯТИЯ ОПТИМИЗАЦИОННЫХ РЕШЕНИЙ, ИНТЕЛЛЕКТУАЛЬНЫЕ АВТОМАТИЗИРОВАННЫЕ ПРОИЗВОДСТВА**.

Содержание дисциплины является основой для освоения дисциплин: **ПОДГОТОВКА К ПРОЦЕДУРЕ ЗАЩИТЫ И ЗАЩИТА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**.

Предварительные компетенции, сформированные у обучающегося до начала изучения дисциплины:

- ОПК-2 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
- ОПК.Д-10 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
- ПК-93 — Способен генерировать новые идеи для решения задач цифровой экономики, абстрагироваться от стандартных моделей, перестраивать сложившиеся способы решения задач, выдвигать альтернативные варианты действий с целью выработки новых оптимальных алгоритмов
- ПК-И1 — владеет технологиями и инструментами искусственного интеллекта для решения задач профессиональной деятельности
- ПК-И2 — способен применять цифровые производственные системы в области профессиональной деятельности

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч.

3.1. Содержание (дидактика) дисциплины

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме		Самостоятельная работа студентов	Формируемая компетенция, %	
				ВСЕГО	Практические занятия		ПК-И1	ПК-И2
4	7	Раздел 1. Введение в обработку больших данных в технических системах. 1.1. Особенности больших данных в авиационных системах: телеметрия, логи, сенсорные потоки 1.2. Архитектуры распределённых систем: Lambda, Kappa 1.3. Обзор инструментов: Hadoop, Spark, Kafka, Flink.	14	4	4	10	15	15
4	7	Раздел 2. Распределённая обработка данных: Apache Spark. 2.1. Архитектура Spark: RDD, DataFrame, Dataset 2.2. Оптимизация вычислений: кэширование, партиционирование 2.3. Практикум: обработка телеметрических данных ЛА в PySpark.	20	8	8	12	20	15
4	7	Раздел 3. Потокковая обработка данных: Apache Kafka и Kafka Streams. 3.1. Принципы event-driven архитектуры 3.2. Настройка Kafka-кластера для приёма данных с датчиков 3.3. Реализация потоковых пайплайнов для мониторинга в реальном времени.	18	6	6	12	20	15
4	7	Раздел 4. Машинное обучение на больших данных. 4.1. Масштабируемые алгоритмы: линейные модели, деревья, ансамбли в Spark MLlib 4.2. Предиктивная аналитика для диагностики систем управления 4.3. Валидация моделей на распределённых данных.	18	6	6	12	20	10
4	7	Раздел 5. Классические статистические методы в аналитике больших данных. 5.1. Описательная статистика и визуализация распределений в больших данных 5.2. Проверка статистических гипотез и дисперсионный анализ (ANOVA) в технических задачах 5.3. Регрессионный анализ и корреляция для выявления зависимостей параметров СУ ЛА 5.4. Статистическое управление качеством (SPC) в производстве.	20	6	6	14	10	25
4	7	Раздел 6. Внедрение и эксплуатация решений Big Data. 6.1. MLOps: версионирование, деплой, мониторинг моделей 6.2. Обеспечение надёжности и безопасности распределённых систем 6.3. Итоговый проект: комплексное решение для обработки данных СУ ЛА.	18	4	4	14	15	20
Всего за 7 семестр			108	34	34	74	100	100
Всего по дисциплине			108	34	34	74	100	100

3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 1. Введение в обработку больших данных в технических системах.	Анализ источников больших данных в авиационных системах. Настройка среды разработки (Docker, Jupyter, PySpark)	2
2		Работа с распределёнными файловыми системами. Загрузка и предобработка телеметрических данных	2
3	Раздел 2. Распределённая обработка данных: Apache Spark.	Практикум: анализ временных рядов телеметрии для выявления аномалий	3
4		Текущий контроль (неделя 6): тестирование по разделам 1–2	1
5		Основы PySpark: RDD и DataFrame API. Агрегация и фильтрация данных телеметрии	2
6		Оптимизация Spark-задач: партиционирование, кэширование, broadcast-переменные	2
7	Раздел 3. Потокковая обработка данных: Apache Kafka и Kafka Streams.	Развёртывание Kafka-кластера в локальной среде. Продюсеры и консьюмеры данных	2
8		Потокковая обработка: Kafka Streams и Structured Streaming. Окна, агрегации, джойны	2
9		Реализация пайплайна мониторинга параметров СУ ЛА в реальном времени	2
10	Раздел 4. Машинное обучение на больших данных.	Введение в Spark MLlib: подготовка данных, пайплайны, кросс-валидация	2
11		Обучение моделей классификации и регрессии на распределённых данных	3
12		Текущий контроль (неделя 10): тестирование по разделам 3–4	1

13	Раздел 5. Классические статистические методы в аналитике больших данных.	Расчёт описательных статистик и построение распределений для параметров СУ ЛА в PySpark/Python	2
14		Проверка статистических гипотез (t-тест, хи-квадрат) для сравнения режимов работы систем	2
15		Построение регрессионных моделей и контрольных карт качества (SPC) для производственных данных	2
16	Раздел 6. Внедрение и эксплуатация решений Big Data.	Основы MLOps: версионирование данных и моделей (DVC, MLflow)	1
17		Защита итоговых проектов. Текущий контроль (неделя 16): тестирование + презентация проект2	3
Всего за 7 семестр			34

3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Введение в обработку больших данных в технических системах.	Установка и настройка Docker-окружения	2
2		Подготовка к практическим занятиям	4
3		Изучение архитектуры распределённых систем по рекомендованной литературе.	4
4	Раздел 2. Распределённая обработка данных: Apache Spark.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	3
5		Выполнение индивидуального практического задания: анализ датасета телеметрии	4
6		Подготовка к текущему контролю (неделя 6): повторение теории, решение типовых задач	2
7		Самостоятельное выполнение упражнений по PySpark.	3
8	Раздел 3. Поточковая обработка данных: Apache Kafka и Kafka Streams.	Изучение документации Apache Kafka.	4
9		Разработка простого потокового пайплайна для обработки синтетических данных датчиков	4
10		Настройка локального кластера, эксперименты с продюсерами/консьюмерами	4
11	Раздел 4. Машинное обучение на больших данных.	Выполнение индивидуального практического задания	4
12		Изучение алгоритмов MLlib. Тренировка моделей на предоставленных данных телеметрии	6
13		Подготовка к текущему контролю (неделя 10): анализ ошибок, повторение ключевых концепций	2
14	Раздел 5. Классические статистические методы в аналитике больших данных.	Подготовка к практическим занятиям	2
15		Изучение методов описательной статистики и визуализации распределений. Анализ гистограмм и квантилей телеметрии	4
16		Выполнение задания по проверке статистических гипотез и дисперсионному анализу на выборках данных испытаний	4
17		Построение регрессионных зависимостей и карт статистического контроля процессов (SPC)	4
18	Раздел 6. Внедрение и эксплуатация решений Big Data.	Изучение основ MLOps.	4
19		Версионирование собственного проекта с использованием DVC/MLflow	2
20		Подготовка и оформление итогового проекта.	8
Всего за 7 семестр			74

4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

СЕМЕСТР	НЕДЕЛИ СЕМЕСТРА																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7		Вопр. Зач	Задан	ИПЗ		ДР		Задан		ДР	Задан	ИПЗ	Контр.Р.	ИПЗ		ДР	ИПЗ, зач.

Условные обозначения:

- ДР – диагностическая работа;
- Вопр. Зач – вопросы к зачету;
- Задан – задание;
- ИПЗ – индивидуальное практическое задание;
- Контр.Р. – контрольная работа;
- зач. – зачет.

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- вопросы к зачету;
- задание;
- индивидуальное практическое задание;
- контрольная работа.

Промежуточная аттестация проводится в формах:

- зачет.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература по дисциплине:

1. А. Бурков. . Машинное обучение без лишних слов. Санкт-Петербург: Питер, 2020, эл. рес.
2. А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкарь. . Большие данные. Big Data. Санкт-Петербург: Лань, 2022, эл. рес.
3. Б. Б. Мойзес, И. В. Плотникова, Л. А. Редько. . Статистические методы контроля качества и обработка экспериментальных данных. Москва: Юрайт, 2022, эл. рес.
4. В. В. Ходосов. . Дисперсионный анализ. Примеры реализации. СПб.: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2023, 16 экз.
5. В. В. Ходосов. . Корреляционный анализ при обработке экспериментальных данных. Санкт-Петербург: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2024, 20 экз.
6. В. С. Михайлов, Н. К. Юрков. . Интегральные оценки в теории надежности. Введение и основные результаты. Москва: Техносфера, 2020, эл. рес.
7. С. Г. Толмачёв. . Нейросетевые методы обработки информации. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2021, 34 экз.

5.2. Дополнительная литература по дисциплине:

не требуется.

5.3. Периодические издания:

1. Информационно-измерительные и управляющие системы.

5.4. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины, электронные библиотечные системы:

1. А.В. Макшанов. Большие данные. Big Data (2024) [Электронный ресурс] : учебник для вузов URL: <https://e.lanbook.com/book/362318>;
2. Горожанина, Е. И. Высокопроизводительные вычисления и анализ больших данных : учебное пособие / Е. И. Горожанина. — Самара : ПГУТИ, 2022. — 132 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/411386>;
3. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/93571>;
4. Ланских, Ю. В. Введение в большие данные : учебное пособие / Ю. В. Ланских, В. Г. Ланских, К. В. Родионов. — Киров : ВятГУ, 2023. — 172 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/408566>;
5. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/93571> - 12 гл.;
6. Гильванов, Р. Г. Надежность информационных систем : учебное пособие / Р. Г. Гильванов, А. В. Забродин. — Санкт-Петербург : ПГУПС, 2022. — 85 с. — ISBN 978-5-7641-1821-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/279020>.

Современные профессиональные базы данных:

1. <https://rusneb.ru> – Национальная электронная библиотека (НЭБ);
2. <https://cyberleninka.ru/> - Научная электронная библиотека «Киберленинка»;
<http://www.rfbr.ru/rffi/ru/library> - Полнотекстовая электронная библиотека Российского фонда фундаментальных исследований.

Информационные справочные системы:

1. Техэксперт – Информационный портал технического регулирования: Нормы, правила, стандарты РФ;

2. http://library.voenmeh.ru/jirbis2/index.php?option=com_irbis&view=irbis&Itemid=457 - БД ГОСТов собственной генерации БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова;
3. <http://www.consultant.ru/> - КонсультантПлюс- информационный портал правовой информации.

5.5. Программное обеспечение:

не требуется.

5.6. Информационные технологии:

взаимодействие с обучающимися посредством ЭИОС Moodle БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Практические занятия:

1. Проектор;
2. Интерактивная доска.

6.2. Прочее:

1. рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
2. рабочие места студентов, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде.

Аннотация рабочей программы

Дисциплина **МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ** является дисциплиной **части, формируемой участниками образовательных отношений блока 1**, программы подготовки по направлению 24.05.06 *Системы управления летательными аппаратами*. Дисциплина реализуется на факультете И Информационные и управляющие системы БГТУ "ВОЕНМЕХ" им. Д.Ф. Устинова кафедрой ИЗ Системы управления и компьютерные технологии.

Дисциплина нацелена на формирование *компетенций*:

ПК-И1 владеет технологиями и инструментами искусственного интеллекта для решения задач профессиональной деятельности;

ПК-И2 способен применять цифровые производственные системы в области профессиональной деятельности.

Содержание дисциплины охватывает круг вопросов, связанных с использованием специализированных методов и инструментов для обработки больших данных; поступающих от производственных систем ракетно-космических комплексов и стрелково-пушечного вооружения, решение задач прогнозной аналитики, оптимизации технологических процессов, повышения надёжности и качества выпускаемой продукции.

Программой дисциплины предусмотрены следующие **виды контроля**:

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- вопросы к зачету;
- задание;
- индивидуальное практическое задание;
- контрольная работа.

Промежуточная аттестация проводится в формах:

- зачет.

Общая трудоемкость освоения дисциплины составляет 3 з.е., **108 ч**. Программой дисциплины предусмотрены практические занятия (**34 ч.**), самостоятельная работа студента (**74 ч.**).

ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 108 ч., из них 34 ч. аудиторных занятий, и 74 ч., отведенных на самостоятельную работу студента.

Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины приведены в таблице.

Контроль освоения дисциплины производится в соответствии с Положением о текущем, рубежном контроле успеваемости и промежуточной аттестации обучающихся.

Формы контроля и критерии оценивания приведены в приложении 3 к Рабочей программе.

Наименование работы	Рекомендуемая литература	Трудоемкость, час.
Раздел 1. Введение в обработку больших данных в технических системах.		
Установка и настройка Docker-окружения	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкарь. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (1,2) С. Г. Толмачёв. . Нейросетевые методы обработки информации: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2021 (1)	2
Подготовка к практическим занятиям		4
Изучение архитектуры распределённых систем по рекомендованной литературе.		4
Итого по разделу 1		10
Раздел 2. Распределённая обработка данных: Apache Spark.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкарь. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (6)	3
Выполнение индивидуального практического задания: анализ датасета телеметрии		4
Подготовка к текущему контролю (неделя 6): повторение теории, решение типовых задач		2
Самостоятельное выполнение упражнений по PySpark.		3
Итого по разделу 2		12
Раздел 3. Поточковая обработка данных: Apache Kafka и Kafka Streams.		
Изучение документации Apache Kafka.	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкарь. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (11-12)	4
Разработка простого потокового пайплайна для обработки синтетических данных датчиков		4
Настройка локального кластера, эксперименты с продюсерами/консьюмерами		4
Итого по разделу 3		12
Раздел 4. Машинное обучение на больших данных.		
Выполнение индивидуального практического задания	А. Бурков. . Машинное обучение без лишних слов: Санкт-Петербург: Питер, 2020 (7)	4
Изучение алгоритмов MLlib. Тренировка моделей на предоставленных данных телеметрии		6
Подготовка к текущему контролю		2

(неделя 10): анализ ошибок, повторение ключевых концепций		
Итого по разделу 4		12
Раздел 5. Классические статистические методы в аналитике больших данных.		
Подготовка к практическим занятиям	Б. Б. Мойзес, И. В. Плотникова, Л. А. Редько. . Статистические методы контроля качества и обработка экспериментальных данных: Москва: Юрайт, 2022 (3,4) В. В. Ходосов. . Дисперсионный анализ. Примеры реализации: СПб.: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2023 (1-4) В. В. Ходосов. . Корреляционный анализ при обработке экспериментальных данных: Санкт-Петербург: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2024 (1-10)	2
Изучение методов описательной статистики и визуализации распределений. Анализ гистограмм и квантилей телеметрии		4
Выполнение задания по проверке статистических гипотез и дисперсионному анализу на выборках данных испытаний		4
Построение регрессионных зависимостей и карт статистического контроля процессов (SPC)		4
Итого по разделу 5		14
Раздел 6. Внедрение и эксплуатация решений Big Data.		
Изучение основ MLOps.	В. С. Михайлов, Н. К. Юрков. . Интегральные оценки в теории надежности. Введение и основные результаты: Москва: Техносфера, 2020 (1-3)	4
Версионирование собственного проекта с использованием DVC/MLflow		2
Подготовка и оформление итогового проекта.		8
Итого по разделу 6		14

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств, позволяющие оценить результаты обучения по данной дисциплине, включают в себя:

- диагностическая работа
- вопросы к зачету;
- задание;
- индивидуальное практическое задание;
- контрольная работа;
- зачет.

Критерии оценивания

Диагностическая работа

Диагностическая работа проводится в форме теста в ЭИОС Moodle:

- при правильном ответе менее чем на 60% вопросов - не аттестация;
- при правильном ответе на 60% вопросов и более - аттестация.

Вопросы к зачету

Вопросы к зачету расположены в УМК дисциплины. Вопросы выдаются преподавателем заранее. При подготовке стоит пользоваться материалом к практическим занятиям, а также источниками основной и дополнительной литературы. При возникновении затруднений студент может обратиться к преподавателю в часы консультаций.

Задание

Задание представлено в срок, не представлен чужой отчет. Каждое задание разбито на 3-5 задач с последовательным увеличением нагрузки для корректного освоения требуемых компетенций. По всем заданиям необходимо успешное выполнение пунктов задания на компьютере, оформление отчета в соответствии с требованиями ГОСТ и успешная защита в установленный срок. Количество баллов и критерии регламентируется Технологической картой дисциплины.

Индивидуальное практическое задание

Индивидуальное практическое задание выполняется на практических занятиях и в часы самостоятельной работы в соответствии с темой, определенной индивидуально для каждого обучающегося.

Практическое задание включает в себя следующие этапы:

1. Постановка цели и задач анализа больших данных.
2. Составление плана исследования в соответствии с предметной областью индивидуального задания. Формулирование гипотез для исследования.
3. Сбор данных и их очистка и систематизация.
4. Первичный анализ и описательная статистика.
5. Проверка гипотез изученными аналитическими методами; подтверждение графическими методами.
6. Написание программы автоматизации этапов анализа.
6. Оформление результатов и выводов.

Результаты выполнения этапов индивидуального практического задания выполняются средствами изученных программных пакетов и демонстрируются преподавателю на практических занятиях.

Контрольная работа

Баллы за контрольную работу проставляются согласно Технологической карте в соответствии с количеством выполненных на практическом занятии заданий средствами изученных программных пакетов.

Зачет

Итоговый контроль по дисциплине проходит в форме зачета.

Зачет считается сданным, если сданы все задания, в соответствии с требованиями, зафиксированными в технологической карте освоения дисциплины (не менее 60 баллов).

Если обучающийся не набрал нужное количество баллов по дисциплине согласно технологической карте, то ему необходимо сдать зачет в очном формате.

Зачет состоит из теоретического вопроса и практической задачи.

Критерии оценивания на очном зачете

Зачтено

1. Дан полный ответ на теоретический вопрос или ответ содержит незначительные ошибки (обучающийся в полной мере не владеет терминологией изученной дисциплины).
2. Задача решена верно или в решении задачи допущены ошибки, которые не приводят к большим отклонениям от правильного ответа, показаны не достаточно прочные практические навыки.

Не зачтено

1. Предварительно не защищены все работы в соответствии с технологической картой.
2. Ответы на вопросы имеют значительные ошибки (неточно указана формула, обучающийся не владеет терминологией изученной дисциплины).
3. Задача решена неверно, допущены грубые ошибки.

Паспорт фонда оценочных средств

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме		Самостоятельная работа студентов	Формируемая компетенция, %		НАИМЕНОВАНИЕ ОЦЕНОЧНОГО СРЕДСТВА
				ВСЕГО	Практические занятия		ПК-И1	ПК-И2	
4	7	Раздел 1. Введение в обработку больших данных в технических системах.	14	4	4	10	15	15	Вопросы к зачету
4	7	Раздел 2. Распределённая обработка данных: Apache Spark.	20	8	8	12	20	15	Задание, Индивидуальное практическое задание
4	7	Раздел 3. Потокковая обработка данных: Apache Kafka и Kafka Streams.	18	6	6	12	20	15	Задание
4	7	Раздел 4. Машинное обучение на больших данных.	18	6	6	12	20	10	Вопросы к зачету, Индивидуальное практическое задание, Задание
4	7	Раздел 5. Классические статистические методы в аналитике больших данных.	20	6	6	14	10	25	Контрольная работа, Индивидуальное практическое задание
4	7	Раздел 6. Внедрение и эксплуатация решений Big Data.	18	4	4	14	15	20	Вопросы к зачету, Индивидуальное практическое задание
Всего за 7 семестр			108	34	34	74	100	100	
Всего по дисциплине			108	34	34	74	100	100	

Оценочные материалы по дисциплине МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ

ПК-И1 - владеет технологиями и инструментами искусственного интеллекта для решения задач профессиональной деятельности

- № 1 Прочитайте текст и запишите развернутый обоснованный ответ
Объясните, почему при обучении модели классификации отказов на несбалансированных данных телеметрии (95% «норма», 5% «отказ») метрика ассигасу может вводить в заблуждение.
Предложите две альтернативные метрики для оценки качества модели.
- № 2 Прочитайте текст и установите соответствие
Распределите компоненты архитектуры Apache Spark в соответствии с их функциональным назначением:

Компонент	Назначение
1. Driver Program	А. Распределённое хранение данных с репликацией
2. Executor	Б. Координация выполнения задач и управление контекстом
3. HDFS	В. Выполнение задач обработки данных на узлах кластера
4. Cluster Manager	Г. Выделение ресурсов для приложений Spark

- № 3 Прочитайте текст и установите соответствие
Соотнесите статистические методы с задачами анализа телеметрических данных СУ ЛА:

Метод	Задача анализа
1. Критерий χ^2 (хи-квадрат)	А. Сравнение средних значений параметра в двух режимах эксплуатации
2. t-критерий Стьюдента	Б. Проверка гипотезы о независимости двух категориальных параметров системы
3. Коэффициент корреляции Пирсона	В. Оценка линейной зависимости между температурой и вибрацией двигателя
4. Доверительный интервал	Г. Интервальная оценка истинного значения параметра с заданной надёжностью

- № 4 Прочитайте текст и установите последовательность
Восстановите порядок операций для системы мониторинга температуры с гарантией *exactly-once*.

- А. Агрегация показаний по 10-секундным окнам в Spark Streaming
- Б. Визуализация динамики в Apache Superset через подключение к ClickHouse
- В. Запись агрегированных данных в ClickHouse
- Г. Генерация JSON-сообщений датчиками в топик Kafka «sensors_temp»
- Д. Настройка контрольных точек (checkpoints) в HDFS для восстановления состояния

- № 5 Прочитайте текст и установите последовательность
Расположите этапы предварительной обработки временных рядов телеметрии СУ ЛА перед подачей данных в нейронную сеть в корректном порядке:

- А. Нормализация или стандартизация числовых признаков (Scaling)
- Б. Обработка пропущенных значений (Imputation) и удаление дубликатов
- В. Генерация признаков (Feature Engineering): скользящие средние, лаги

Г. Синхронизация временных меток и ресемплинг к единому шагу времени

Д. Разделение выборки на обучающую, валидационную и тестовую (с учётом времени)

№ 6 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Какой метод предобработки данных в PySpark наиболее эффективен для обработки пропущенных значений в телеметрии при сохранении распределения признака?

1. Удаление всех строк с пропусками (dropna)
2. Замена пропусков на среднее значение столбца
3. Импутация с использованием модели KNN или медианы по группам
4. Замена пропусков на нулевое значение

№ 7 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Для обеспечения отказоустойчивости потоковой обработки данных в Apache Kafka при обработке телеметрии в реальном времени необходимо:

1. Использовать один раздел (partition) для всех сообщений
2. Настроить репликацию топиков и acks=all для продюсеров
3. Отключить сохранение сообщений после обработки
4. Обрабатывать данные только на одном потребительском узле

№ 8 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

При обработке телеметрических данных СУ ЛА в PySpark необходимо исключить из анализа записи, где параметр «скорость вращения турбины» выходит за физически допустимые пределы (0–25000 об/мин). Какой метод фильтрации обеспечит корректную обработку с учётом возможных пропущенных значений (null)?

1. `df.filter(col("rpm") >= 0) & (col("rpm") <= 25000)`
2. `df.filter(col("rpm").between(0, 25000) & col("rpm").isNotNull())`
3. `df.na.drop("rpm").filter(col("rpm") < 25000)`
4. `df.where(col("rpm") != null).filter(col("rpm") <= 25000)`

№ 9 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Какие характеристики данных телеметрии указывают на необходимость применения распределённой обработки вместо локального анализа?

1. Объём данных превышает доступную оперативную память одного узла
2. Данные поступают с частотой более 1000 сообщений в секунду
3. Требуется обработка с задержкой не более 100 мс
4. Все данные имеют одинаковый тип и формат

№ 10 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Отметьте все операции, которые можно выполнить в PySpark для выявления аномалий в параметрах телеметрии СУ ЛА:

1. Расчёт z-score для каждого значения параметра
2. Применение межквартильного размаха (IQR) для фильтрации выбросов
3. Использование изолирующего леса (Isolation Forest) из Spark MLlib
4. Визуализация данных в виде круговой диаграммы

№ 11 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Отметьте все признаки, указывающие на необходимость применения оконных функций (window functions) в Spark для анализа телеметрии СУ ЛА:

1. Требуется расчёт скользящего среднего параметра за последние N измерений

2. Необходимо ранжировать события по времени внутри каждой группы «идентификатор полёта»
3. Требуется агрегировать данные по всем записям датасета без учёта временного порядка
4. Необходимо вычислить накопительную сумму параметра с начала полёта для каждого борта

№ 12 Прочитайте текст и запишите развернутый обоснованный ответ

Сформулируйте условие фильтрации в PySpark для выбора записей телеметрии, где параметр «давление» находится в диапазоне 200–220 бар И параметр «температура» не превышает 85°C, исключая записи с флагом «sensor_error = true».

ПК-И2 - способен применять цифровые производственные системы в области профессиональной деятельности

№ 1 Прочитайте текст и установите последовательность

Установите корректную последовательность этапов анализа распределения параметра телеметрии СУ ЛА с использованием описательной статистики и визуализации.

А. Построение гистограммы и box-plot для визуальной оценки формы распределения и выбросов

Б. Расчёт описательных статистик: среднее, медиана, стандартное отклонение, квантили

В. Проверка гипотезы о нормальности распределения (критерий Шапиро-Уилка или Колмогорова-Смирнова)

Г. Очистка данных: обработка пропусков и детекция аномальных значений на основе IQR или z-score

Д. Интерпретация результатов: выбор параметрических или непараметрических методов для дальнейшего анализа

№ 2 Прочитайте текст и запишите развернутый обоснованный ответ

Обоснуйте выбор между линейной регрессией и множественной регрессией для моделирования зависимости параметра «точность стабилизации курса» от факторов: скорость полёта, высота, температура двигателя, уровень вибрации.

№ 3 Прочитайте текст и запишите развернутый обоснованный ответ

Среднее по выборке измерений признака составило 23,5.

Среднеквадратическое отклонение равно 0,15.

Необходимо проверить один сомнительный результат измерения, равный 24,4.

Выберите критерий проверки, не требующий использование вспомогательных таблиц, рассчитайте необходимое значение критерия и сделайте вывод.

№ 4 Прочитайте текст и установите соответствие

Соотнесите протоколы обмена данными с областями их применения в цифровых производственных системах авиационной отрасли:

Протокол	Область применения
1. MQTT	А. Лёгкий протокол для передачи телеметрии с IoT-датчиков
2. OPC UA	Б. Промышленный стандарт для интеграции оборудования и SCADA-систем
3. REST API	В. Веб-интерфейсы для доступа к аналитическим сервисам
4. Modbus TCP	Г. Обмен данными с программируемыми логическими контроллерами (ПЛК)

№ 5 Прочитайте текст и установите соответствие

Распределите компоненты цифрового двойника системы управления ЛА в соответствии с их функциями:

Компонент

Функция

Компонент	Функция
1. Data Ingestion Layer	А. Выполнение моделей прогнозирования и оптимизации
2. Analytics Engine	Б. Приём и буферизация потоковых данных от физических систем
3. Visualization Dashboard	В. Представление KPI и алертов для операторов
4. Feedback Loop	Г. Передача управляющих воздействий обратно в физическую систему

№ 6 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Отметьте все требования, которые должны выполняться при интеграции аналитического модуля Big Data в контур управления качеством авиационной продукции:

1. Соответствие отраслевым стандартам верификации и валидации (DO-178C, DO-254 при необходимости)
2. Наличие механизмов объяснимости моделей ИИ для прохождения сертификационных процедур
3. Полная автономность модуля от внешних систем для повышения безопасности
4. Возможность трассировки входных данных к исходным источникам (датчикам, журналам испытаний)

№ 7 Прочитайте текст и установите последовательность

Расположите шаги проведения дисперсионного анализа (ANOVA) для сравнения эффективности трёх режимов работы системы управления ЛА.

А. Проверка допущений ANOVA: нормальность остатков, гомогенность дисперсий (тест Левена), независимость наблюдений

Б. Формулировка нулевой и альтернативной гипотез: H_0 — средние равны во всех группах, H_1 — хотя бы одно среднее отличается

В. Проведение пост-хок тестов (Тьюки, Бонферрони) для попарного сравнения групп при отклонении H_0

Г. Расчёт F-статистики и p-value, принятие решения об отклонении или непринятии нулевой гипотезы

Д. Подготовка данных: группировка наблюдений по фактору «режим работы», проверка баланса выборок

№ 8 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Для обеспечения прослеживаемости данных в цифровой производственной системе при расследовании инцидента с системой управления ЛА критически важным является:

1. Использование сжатия данных для экономии дискового пространства
2. Ведение неизменяемого журнала аудита (immutable audit log) с фиксацией времени и источника каждого изменения
3. Хранение данных только в оперативной памяти для ускорения доступа
4. Автоматическое удаление «устаревших» записей старше 24 часов

№ 9 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Для обеспечения целостности данных при интеграции испытательного стенда с цифровой производственной системой наиболее критичным является:

1. Использование транзакций и механизмов подтверждения доставки сообщений
2. Хранение данных только в формате JSON
3. Отключение логирования для повышения производительности
4. Передача данных без шифрования для ускорения обработки

№ 10 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Какой подход к статистическому управлению качеством (SPC) позволяет оперативно выявлять разладку технологического процесса изготовления узлов СУ ЛА?

1. Построение гистограммы по итогам месяца
2. Использование контрольных карт Шухарта с правилами Вестерн Электрик
3. Расчёт среднего значения за весь период производства
4. Визуальный осмотр готовой продукции

№ 11 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Какие требования должны выполняться для корректной интеграции данных от разнородных источников в единую цифровую производственную систему?

1. Единый формат временных меток (например, UTC)
2. Синхронизация часов всех источников с точностью до миллисекунд
3. Использование одинаковых наименований параметров во всех системах
4. Наличие метаданных, описывающих семантику и единицы измерения

№ 12 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Отметьте все утверждения, корректно описывающие применение статистических методов для анализа телеметрии СУ ЛА.

1. Для визуализации распределения параметра «давление в гидросистеме» при объёме выборки $>10\,000$ наблюдений целесообразно использовать гистограмму с автоматическим подбором числа бинов (правило Стерджеса или Фридмана-Дьякониса)
2. При проверке гипотезы о равенстве дисперсий двух выборок телеметрии критерий Фишера (F-тест) применим только при условии нормальности распределения исходных данных
3. Коэффициент корреляции Пирсона позволяет выявить нелинейные зависимости между параметрами «температура двигателя» и «расход топлива»
4. В множественной регрессии увеличение Adjusted R^2 при добавлении нового предиктора свидетельствует об улучшении качества модели с учётом штрафа за усложнение
5. Для анализа влияния категориального фактора «тип полётного режима» на непрерывный параметр «точность навигации» при трёх и более уровнях фактора оптимален однофакторный ANOVA