

УТВЕРЖДАЮ  
 Декан факультета

« \_\_\_\_ » \_\_\_\_\_ 20\_\_

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ

Направление/специальность подготовки	24.05.02 Проектирование авиационных и ракетных двигателей
Специализация/профиль/программа подготовки	Проектирование технологических процессов производства авиационных, ракетных двигателей и энергетических установок
Уровень высшего образования	Специалитет
Форма обучения	Очная
Факультет	А Ракетно-космическая техника
Выпускающая кафедра	А8 ДВИГАТЕЛИ И ЭНЕРГОУСТАНОВКИ ЛЕТАТЕЛЬНЫХ АППАРАТОВ
Кафедра-разработчик рабочей программы	Н2 Программная инженерия и интеллектуальные системы

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	
2	3	2	72	34	17	0	17	38	0	0	38	диф. зач.

*ЛИСТ СОГЛАСОВАНИЯ*

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

**24.05.02 Проектирование авиационных и ракетных двигателей**

год набора группы: 2026

Программу составил:

Кафедра Н2 Программная инженерия и интеллектуальные системы  
Ярошевская Елена Юрьевна, старший преподаватель

\_\_\_\_\_

Программа рассмотрена  
на заседании кафедры-разработчика  
рабочей программы **Н2 Программная инженерия и интеллектуальные системы**

Заведующий кафедрой Семенова Е.Г., д.т.н., проф.

\_\_\_\_\_

Программа рассмотрена  
на заседании выпускающей кафедры

**А8 ДВИГАТЕЛИ И ЭНЕРГОУСТАНОВКИ ЛЕТАТЕЛЬНЫХ АППАРАТОВ**

Заведующий кафедрой Саваровский А.А., к.т.н.

\_\_\_\_\_

# **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

## **ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ**

### **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-94 — Способен к управлению информацией и данными, поиску источников информации и данных, восприятию, анализу, запоминанию и передаче информации с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач

Формированию компетенций служит достижение следующих результатов образования:

### **ПК-94**

*знания:*

основных концепций, характеристик (5V) и архитектурных подходов к работе с большими данными;

типичных источников инженерных данных: телеметрия стендовых и лётных испытаний, материалы лабораторных исследований, данные эксплуатации, а также методов их первичной классификации и оценки релевантности;

базовых алгоритмов и цифровых инструментов (Python, pandas, основы Spark) для сбора, очистки, агрегации и визуализации данных, необходимых для подготовки информации к анализу;

*умения:*

осуществлять целенаправленный поиск и критическую оценку источников данных;

применять алгоритмы предварительной обработки данных: фильтрация шумов, обработка пропущенных значений, нормализация и агрегация временных рядов телеметрии;

использовать цифровые средства визуализации для представления результатов анализа в форме, обеспечивающей эффективное восприятие и передачу информации при принятии инженерных решений;

*навыки:*

практического применения скриптов на Python/pandas для загрузки и первичного анализа структурированных и полуструктурированных данных из разнородных источников;

документирования процессов сбора и обработки данных в соответствии с требованиями воспроизводимости исследований и командной работы;

передачи результатов анализа информации с использованием интерактивных дашбордов и отчётных форм для поддержки коллективного решения задач проектирования.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению *24.05.02 Проектирование авиационных и ракетных двигателей*.

Содержание дисциплины является логическим продолжением дисциплин: **ПРОГРАММНАЯ ИНЖЕНЕРИЯ: АЛГОРИТМИЗАЦИЯ И ОБРАБОТКА ДАННЫХ**.

Содержание дисциплины является основой для освоения дисциплин: **УПРАВЛЕНИЕ БОЛЬШИМИ ДАННЫМИ**.

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
- ПК-91 — Способен к коммуникации и кооперации в цифровой среде, использованию различных цифровых средств, позволяющих во взаимодействии с другими людьми достигать поставленных целей
- ПСК-5.12 — Способен применять современные языки программирования при решении задач профессиональной деятельности

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 2 з.е., 72 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	Лекции	Практические занятия		ПК-94
2	3	<b>Раздел 1. Основы концепции больших данных в инженерной практике.</b> 1.1. Характеристики 5V, классификация данных, жизненный цикл информации. 1.2. Особенности больших данных в авиационных системах: телеметрия, логи, сенсорные потоки. 1.3. Архитектуры распределённых систем: Lambda, Kappa.	9	5	4	1	4	20
2	3	<b>Раздел 2. Источники и методы первичного сбора данных.</b> 2.1. Типология источников инженерных данных (датчики, журналы испытаний, CAD-модели, нормативные базы). 2.2. Методы поиска, оценки релевантности и первичной классификации данных. 2.3. Основы работы с открытыми и корпоративными репозиториями технической информации.	10	6	4	2	4	20
2	3	<b>Раздел 3. Алгоритмы предварительной обработки и очистки данных.</b> 3.1. Выявление и обработка пропущенных значений, фильтрация шумов в телеметрических потоках, нормализация и агрегация временных рядов. 3.2. Базовые алгоритмы трансформации данных для обеспечения качества входной информации. 3.3. Описательная статистика больших данных.	10	6	4	2	4	20
2	3	<b>Раздел 4. Инструменты визуализации и интерпретации результатов.</b> 4.1. Содержание: принципы эффективного представления инженерных данных. 4.2. Цифровые средства визуализации (Matplotlib, Plotly, базовые дашборды) для передачи аналитических выводов при принятии проектных решений.	10	6	4	2	4	20
2	3	<b>Раздел 5. Практикум: анализ учебного набора данных.</b> 5.1. Выполнение проекта по загрузке, очистке, анализу и визуализации репрезентативного набора данных, имитирующего параметры работы авиационного двигателя. 5.2. Документирование процесса и презентация результатов в формате инженерного отчёта.	33	11	1	10	22	20
<b>Всего за 3 семестр</b>			72	34	17	17	38	100
<b>Всего по дисциплине</b>			72	34	17	17	38	100

#### 3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 1. Основы концепции больших данных в инженерной практике.	Настройка среды разработки (Docker, Jupyter, PySpark)	1
2	Раздел 2. Источники и методы первичного сбора данных.	Анализ источников больших данных в авиационных системах.	2
3	Раздел 3. Алгоритмы предварительной обработки и очистки данных.	Методы нормализации, кодирования и агрегации данных: обеспечение качества входной информации для анализа	2
4	Раздел 4. Инструменты визуализации и интерпретации результатов.	Разработка дашборда как средства визуализации инженерных данных	2
5	Раздел 5. Практикум: анализ учебного набора данных.	Загрузка и первичный разведочный анализ (EDA) телеметрических данных авиационного двигателя: оценка качества и репрезентативности набора данных.	2
6		Очистка и предобработка данных: обработка пропусков, выбросов и аномалий в параметрах работы двигателя с использованием Pandas и Scikit-learn.	2
7		Исследовательский анализ и визуализация ключевых метрик: выявление закономерностей, корреляций и признаков деградации двигателя с помощью Matplotlib и Plotly.	2
8		Разработка интерактивного дашборда для мониторинга состояния двигателя: интеграция визуализаций и	2

		проектирование пользовательского интерфейса.	
9		Подготовка инженерного отчёта: структурирование результатов, документирование пайплайна обработки данных и презентация выводов в формате технического отчёта.	2
<b>Всего за 3 семестр</b>			<b>17</b>

### 3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Основы концепции больших данных в инженерной практике.	Изучение архитектуры распределённых систем по рекомендованной литературе.	2
2		Подготовка к практическим занятиям	2
3	Раздел 2. Источники и методы первичного сбора данных.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	2
4		Установка и настройка Docker-окружения	1
5		Подготовка к практическим занятиям	1
6	Раздел 3. Алгоритмы предварительной обработки и очистки данных.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	1
7		Реализация и сравнительный анализ алгоритмов предварительной обработки и очистки данных на примере реального набора данных	2
8		Подготовка к практическим занятиям	1
9	Раздел 4. Инструменты визуализации и интерпретации результатов.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе <sup>1</sup>	1
10		Сравнительный анализ ПО Matplotlib и Plotly	2
11		Подготовка к практическим занятиям	1
12	Раздел 5. Практикум: анализ учебного набора данных.	Выбор темы. Загрузка и первичная валидация набора данных: проверка целостности, описание схемы данных и формирование метаданных индивидуального задания	3
13		Автоматизация очистки данных: разработка функций для обработки пропущенных значений, дубликатов и некорректных форматов в параметрах двигателя	2
14		Статистический анализ распределений ключевых параметров: расчёт описательных статистик, выявление выбросов методами IQR и Z-score	4
15		Визуализация временных рядов и корреляционных зависимостей: построение графиков деградации параметров и матрицы корреляций с аннотированием выводов	3
16		Прототипирование интерактивной визуализации: создание базовых компонентов дашборда в Plotly Express для исследования сценариев отказа двигателя	3
17		Документирование пайплайна обработки данных: написание технических комментариев, оформление Jupyter Notebook и подготовка репроизводимого кода	3
18		Структурирование инженерного отчёта: формулировка целей, описание методологии, визуальное оформление результатов и подготовка презентации выводов	4
Всего за 3 семестр			38

## 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
3	Вопр.Диф.Зач	Задан	Вопр.Диф.Зач	Задан	ДР	Задан			ДР						ИПЗ	ДР	Вопр.Диф.Зач, диф. зач.

Условные обозначения:

- ДР – диагностическая работа;
- Задан – задание;
- Вопр.Диф.Зач – вопросы к дифференцированному зачету;
- ИПЗ – индивидуальное практическое задание;
- диф. зач. – дифференцированный зачет.

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

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

**Промежуточная аттестация** проводится в формах:

- дифференцированный 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. В. В. Ходосов. . Корреляционный анализ при обработке экспериментальных данных. Санкт-Петербург: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2024, 20 экз.
4. К. Н. Волков, В. И. Запрягаев, В. Н. Емельянов. . Визуализация данных физического и математического моделирования в газовой динамике. М.: ФИЗМАТЛИТ, 2018, 6 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. 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>;;
2. 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>;;
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. 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.,.

### 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](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. Лекционные занятия:**

специализированные требования по оборудованию отсутствуют; аудитория с посадочными местами по количеству студентов; доска.

### **6.2. Практические занятия:**

1. Проектор;
2. Интерактивная доска.

### **6.3. 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 **ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению 24.05.02 *Проектирование авиационных и ракетных двигателей*. Дисциплина реализуется на факультете Н Робототехника и инновационная инженерия БГТУ "ВОЕНМЕХ" им. Д.Ф. Устинова кафедрой Н2 Программная инженерия и интеллектуальные системы.

Дисциплина нацелена на формирование *компетенций*:

ПК-94 Способен к управлению информацией и данными, поиску источников информации и данных, восприятию, анализу, запоминанию и передаче информации с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач.

Содержание дисциплины охватывает круг вопросов, связанных с фундаментальными принципами работы с большими данными, их ключевыми характеристиками, типичными источниками и методами первичной обработки. Особое внимание уделяется специфике применения данных на примерах проектирования авиационных двигателей и энергетических установок: анализа телеметрии, результатов стендовых и лётных испытаний, материаловедческих исследований и эксплуатационной статистики. Студенты осваивают базовые инструменты сбора, очистки, агрегации и визуализации больших данных, формируя исходные компетенции для последующего перехода к системному управлению и аналитике в инженерной практике.

Программой дисциплины предусмотрены следующие **виды контроля**:

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

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

**Промежуточная аттестация** проводится в формах:

- дифференцированный зачет.

Общая трудоемкость освоения дисциплины составляет 2 з.е., 72 ч. Программой дисциплины предусмотрены лекционные занятия (17 ч.), практические занятия (17 ч.), самостоятельная работа студента (38 ч).

## ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

### Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 72 ч., из них 34 ч. аудиторных занятий, и 38 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. Основы концепции больших данных в инженерной практике.		
Изучение архитектуры распределённых систем по рекомендованной литературе.	А. Бурков. . Машинное обучение без лишних слов: Санкт-Петербург: Питер, 2020 (1)	2
Подготовка к практическим занятиям	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкаръ. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (1)	2
Итого по разделу 1		4
Раздел 2. Источники и методы первичного сбора данных.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкаръ. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (2)	2
Установка и настройка Docker-окружения		1
Подготовка к практическим занятиям		1
Итого по разделу 2		4
Раздел 3. Алгоритмы предварительной обработки и очистки данных.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. В. Макшанов, А. Е. Журавлёв, Л. Н. Тындыкаръ. . Большие данные. Big Data: Санкт-Петербург: Лань, 2022 (3)	1
Реализация и сравнительный анализ алгоритмов предварительной обработки и очистки данных на примере реального набора данных		2
Подготовка к практическим занятиям		1
Итого по разделу 3		4
Раздел 4. Инструменты визуализации и интерпретации результатов.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе1	К. Н. Волков, В. И. Запрягаев, В. Н. Емельянов. . Визуализация данных физического и математического моделирования в газовой динамике: М.: ФИЗМАТЛИТ, 2018 (1--3)	1
Сравнительный анализ ПО Matplotlib и Plotly		2
Подготовка к практическим занятиям		1
Итого по разделу 4		4
Раздел 5. Практикум: анализ учебного набора данных.		
Выбор темы. Загрузка и первичная валидация набора данных: проверка целостности, описание схемы данных и формирование метаданных индивидуального задания	А. Бурков. . Машинное обучение без лишних слов: Санкт-Петербург: Питер, 2020 (3-5) В. В. Ходосов. . Корреляционный анализ при обработке экспериментальных данных:	3

Автоматизация очистки данных: разработка функций для обработки пропущенных значений, дубликатов и некорректных форматов в параметрах двигателя	Санкт-Петербург: Изд-во БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2024 (1-2)	2
Статистический анализ распределений ключевых параметров: расчёт описательных статистик, выявление выбросов методами IQR и Z-score		4
Визуализация временных рядов и корреляционных зависимостей: построение графиков деградации параметров и матрицы корреляций с аннотированием выводов		3
Прототипирование интерактивной визуализации: создание базовых компонентов дашборда в Plotly Express для исследования сценариев отказа двигателя		3
Документирование пайплайна обработки данных: написание технических комментариев, оформление Jupyter Notebook и подготовка репроизводимого кода		3
Структурирование инженерного отчёта: формулировка целей, описание методологии, визуальное оформление результатов и подготовка презентации выводов		4
Итого по разделу 5		22

## ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

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

- диагностическая работа
- вопросы к дифференцированному зачету;
- задание;
- индивидуальное практическое задание;
- дифференцированный 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 дифференцированному зачету

Вопросы к дифференцированному зачету расположены в УМК дисциплины. Вопросы выдаются преподавателем заранее. При подготовке стоит пользоваться лекционным материалом, 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 требуемой компетенции. 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. Формулирования 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.

#### Дифференцированный зачет

Итоговый контроль по дисциплине проходит в форме дифференцированного зачета.

Дифференцированный 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 или хочет повысить оценку по дисциплине согласно технологической карте, то ему необходимо сдать Дифференцированный зачет. Дифференцированный зачет состоит из теоретического вопроса и практической задачи.

Критерии оценивания на Дифференцированном зачете .

Оценка «отлично»

1. Защищены все работы в соответствии с технологической картой.

2. Даны полные ответы на вопросы (точно указаны определения, формулы, студент владеет терминологией изученной дисциплины).
3. Правильно решена задача, показано умение грамотно применять полученные теоретические знания в практических целях.

Оценка «хорошо»

1. Предварительно (в установленные сроки) защищены работы все работы в соответствии с технологической картой.
2. Данные ответы на вопросы имеют незначительные ошибки.
3. Правильно решены задачи, но ход их решения не является оптимальным, показаны прочные практические навыки.

Оценка «удовлетворительно»

1. Работы в соответствии с технологической картой защищались с нарушением сроков сдачи.
2. Данные ответы на вопросы имеют незначительные ошибки (обучающийся в полной мере не владеет терминологией изученной дисциплины).
3. В решении задачи допущены ошибки, которые не приводят к большим отклонениям от правильного ответа, показаны не достаточно прочные практические навыки.

Оценка «неудовлетворительно»

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.

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %	НАИМЕНОВАНИЕ ОЦЕНОЧНОГО СРЕДСТВА
				ВСЕГО	Лекции	Практические занятия		ПК-94	
2	3	Раздел 1. Основы концепции больших данных в инженерной практике.	9	5	4	1	4	20	Вопросы к дифференцированному зачету
2	3	Раздел 2. Источники и методы первичного сбора данных.	10	6	4	2	4	20	Задание
2	3	Раздел 3. Алгоритмы предварительной обработки и очистки данных.	10	6	4	2	4	20	Вопросы к дифференцированному зачету, Задание
2	3	Раздел 4. Инструменты визуализации и интерпретации результатов.	10	6	4	2	4	20	Вопросы к дифференцированному зачету, Задание
2	3	Раздел 5. Практикум: анализ учебного набора данных.	33	11	1	10	22	20	Индивидуальное практическое задание
Всего за 3 семестр			72	34	17	17	38	100	
Всего по дисциплине			72	34	17	17	38	100	



## Оценочные материалы по дисциплине ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ

**ПК-94 - Способен к управлению информацией и данными, поиску источников информации и данных, восприятию, анализу, запоминанию и передаче информации с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач**

- № 1 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».
- № 2 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.
- № 3 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
- № 4 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
- № 5 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
- № 6 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
- № 7 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)

№ 8 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

№ 9 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)`

№ 10 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

№ 11 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
	Г. Выделение ресурсов для приложений Spark

№ 12 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. Передача управляющих воздействий обратно в физическую систему Г. Представление KPI и алертов для операторов