

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Хоружий Дарьяна Ивановна

Должность: Директор института экономики и управления АПК

Дата подписания: 14.01.2026 15:01:48

Уникальный идентификатор ключа:

1e90b132d9b04dce67585160b015dddf2cb1e6a9



**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –

МСХА имени К.А. ТИМИРЯЗЕВА»

(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)

Институт экономики и управления АПК
Кафедра статистики и кибернетики

УТВЕРЖДАЮ:

**Директор института
экономики и управления АПК**

Д.И. Хоружий



«28» августа 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.13 Инжиниринг данных

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»

Направленность: «Большие данные и машинное обучение»

Курс 3

Семестр 6

Форма обучения: очная

Год начала подготовки: 2025

Москва, 2025

Разработчики:

Уколова А.В., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)


(подпись)

«26» августа 2025 г.

Бодур А.М., ассистент
(ФИО, ученая степень, ученое звание)


(подпись)

«26» августа 2025 г.

Рецензент:

Вахрушева И.А., канд. пед. наук, доцент
(ФИО, ученая степень, ученое звание)


(подпись)

«26» августа 2025 г.

Программа составлена в соответствии с требованиями ФГОС ВО, компетентностно-ролевых моделей в сфере искусственного интеллекта по направлению подготовки 09.03.02 «Информационные системы и технологии», профессионального стандарта и учебного плана

Программа обсуждена на заседании кафедры статистики и кибернетики. Протокол № 11 от «26» августа 2025 г.

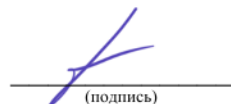
И. о. зав. кафедрой Уколова А.В., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)


(подпись)

«26» августа 2025 г.

Согласовано:

Председатель учебно-методической
комиссии института экономики и управления АПК
Гупалова Т.Н., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)


(подпись)

Протокол №1 «28» августа 2025 г.

И. о. зав. выпускающей кафедрой
статистики и кибернетики
Уколова А.В., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)


(подпись)

«28» августа 2025 г.

Заведующий отделом комплектования ЦНБ


(подпись)
(подпись)

СОДЕРЖАНИЕ

АННОТАЦИЯ	4
1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	5
2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ.....	5
3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	5
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	8
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ ПО СЕМЕСТРАМ	8
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	8
4.3 ЛЕКЦИИ/ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	10
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....	15
6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	16
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ.....	16
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ	21
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	22
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	22
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	23
7.3 СТАТЬИ, ОПУБЛИКОВАННЫЕ В НАУЧНЫХ ЖУРНАЛАХ I УРОВНЯ БЕЛОГО СПИСКА НАУЧНЫХ ЖУРНАЛОВ МИНОБРНАУКИ РОССИИ И СБОРНИКАХ НАУЧНЫХ РАБОТ КОНФЕРЕНЦИЙ УРОВНЯ А*	23
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	23
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ	24
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ	25
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ..	26
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	27

АННОТАЦИЯ

рабочей программы учебной дисциплины Б1.В.13 «Инжиниринг данных» для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии», направленности «Большие данные и машинное обучение»

Цель освоения дисциплины. Основная цель дисциплины «Инжиниринг данных» – овладение студентами основными методами проектирования, создания и сопровождения масштабируемых, надежных и эффективных систем, которые обеспечивают превращение необработанных данных в качественную, пригодную для анализа информацию, тем самым закладывая технологический фундамент для Data Science, машинного обучения и бизнес-аналитики.

Место дисциплины в учебном плане: дисциплина включена в часть учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии», формируемую участниками образовательных отношений.

Требования к результатам освоения дисциплины: в результате освоения дисциплины формируются следующие компетенции (индикаторы):

ПКкрмии-10 (BD-1. Экспертный уровень).1; ПКкрмии-11 (BD-2. Экспертный уровень).1; ПКкрмии-11 (BD-2. Экспертный уровень).2; ПКкрмии-14 (ML-2. Экспертный уровень).2

Краткое содержание дисциплины: Введение в инжиниринг данных. Роль в Data-экосистеме. Основы работы с данными и СУБД. Проектирование и построение ETL/ELT процессов. Извлечение данных из различных источников. Хранилища данных и Data Lakes. Введение в экосистему Hadoop и Apache Spark. Оркестрация пайплайнов данных. Качество данных, мониторинг и DevOps для данных.

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 часа, в т. ч. 4 часа практической подготовки).

Промежуточный контроль: экзамен.

1. Цель освоения дисциплины

Основная цель дисциплины «Инжиниринг данных» – овладение студентами основными методами проектирования, создания и сопровождения масштабируемых, надежных и эффективных систем, которые обеспечивают превращение необработанных данных в качественную, пригодную для анализа информацию, тем самым закладывая технологический фундамент для Data Science, машинного обучения и бизнес-аналитики.

2. Место дисциплины в учебном процессе

Дисциплина «Инжиниринг данных» относится к части Блока 1 «Дисциплины (модули)» учебного плана, формируемой участниками образовательных отношений. Дисциплина «Инжиниринг данных» реализуется в соответствии с требованиями ФГОС, ОПОП ВО, Учебного плана, по направлению 09.03.02 «Информационные системы и технологии» и компетентностно-ролевых моделей в сфере искусственного интеллекта (КРМ ИИ), разработанной по заказу Аналитического центра при Правительстве Российской Федерации в рамках федерального проекта «Искусственный интеллект» национального проекта «Экономика данных и цифровая трансформация государства».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Инжиниринг данных» являются «Математическая статистика», «Теория информации», «Алгоритмизация и программирование», «Программирование на языке Python», «Основы науки о данных (Data Science)» и др.

Дисциплина «Инжиниринг данных» является основополагающей для изучения следующих дисциплин: «Методы искусственного интеллекта», «Инжиниринг данных», «Большие данные в сельском хозяйстве» и др.

Рабочая программа дисциплины «Методы машинного обучения» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Образовательные результаты освоения дисциплины обучающимся, представлены в таблице 1.

Таблица 1

Требования к результатам освоения учебной дисциплины

п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				Знать	Уметь	Владеть
1	ПККрмии-10 (BD-1. Экспертный уровень)	Способен осуществлять поиск, сбор, очистку и предварительный анализ данных	ПККрмии-10 (BD-1. Экспертный уровень).1 Индикатор: Обосновывает способы и варианты применения методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи. Уровень освоения индикатора: Формулирует гипотезы относительно данных, подтверждает или опровергает их	Математические основы и алгоритмическую сущность методов предварительного анализа данных (EDA), принципы их выбора для различных типов задач искусственного интеллекта, а также методологию формирования и проверки статистических гипотез относительно структуры, распределений и взаимосвязей в данных	Применять методы разведочного анализа для проверки сформулированных гипотез о данных, обосновывать выбор конкретных алгоритмов и статистических тестов для их подтверждения или опровержения, а также адаптировать и преобразовывать эти методы под специфику задачи ИИ, интерпретируя полученные результаты	Навыками практического применения инструментов EDA (например, библиотек Python: Pandas, Seaborn, SciPy) для целенаправленной проверки гипотез о данных, включая визуализацию, расчет статистических критериев и алгоритмическую обработку, позволяющую делать обоснованные выводы о пригодности данных для построения моделей машинного обучения
2	ПККрмии-11 (BD-2. Экспертный уровень)	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных,	ПККрмии-11 (BD-2. Экспертный уровень).1 Индикатор: Определяет требования к наборам и качеству данных для решения задач машинного обучения. Уровень освоения индикатора: Разрабатывает требования	Методологию и критерии формирования требований к данным для задач машинного обучения, включая принципы проектирования спецификаций для разметки данных,	Анализировать бизнес-задачу и целевые метрики модели для разработки детальных технических требований к исходным и размеченным данным, а также к последовательности их обработки, обеспечивающей	Практическими навыками создания документов спецификации требований к данным и инструкций по разметке, включая определение форматов, схемы разметки, метрик качества разметки и

		оценивать качество данных, обеспечивать непрерывную интеграцию данных	для разметки и обработки данных	параметров качества и этапов предобработки	соответствие данных нуждам алгоритмов машинного обучения	пайплайнов предварительной обработки данных.
			ПККрмии-11 (BD-2. Экспертный уровень).2 Индикатор: Работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность. Уровень освоения индикатора: Разрабатывает требования для инструментария разметки, оценивает качество данных	Принципы проектирования требований к инструментальным средствам для сбора и разметки данных, включая критерии оценки их функциональной пригодности, а также методологию проведения аудита качества данных на всех этапах жизненного цикла	Формулировать функциональные и технические требования к выбору или разработке инструментария для разметки данных, а также систематически оценивать качество исходных и производных данных, используя стандартные метрики	Навыками разработки технического задания на инструмент разметки, проведения сравнительного анализа существующих платформ и практическими методами верификации качества данных
3	ПККрмии-14 (ML-2. Экспертный уровень)	Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками	ПККрмии-14 (ML-2. Экспертный уровень).2 Индикатор: Применяет методы предварительной обработки данных и работы с признаками. Уровень освоения индикатора: Проектирует и внедряет комплексные пайплайны предварительной обработки данных с использованием современных методов ИИ, автоматизации и feature engineering в различных предметных областях	Теоретические основы и современные методы предобработки данных, feature engineering, а также принципы проектирования сквозных, воспроизводимых и масштабируемых пайплайнов данных с использованием передовых практик автоматизации и инструментов MLOPs.	Проектировать, разрабатывать и внедрять комплексные сквозные пайплайны (ETL/ELT пайплайны) предобработки данных и генерации признаков, адаптированные под специфику различных предметных областей, с использованием современных фреймворков и библиотек для автоматизации и ИИ.	Навыками полного цикла создания пайплайнов данных – от проектирования и кодирования до внедрения, используя инструменты для автоматизации обработки, feature engineering и эффективного управления признаковым пространством в задачах ИИ.

4. Структура и содержание дисциплины

4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 4 зачетные единицы (144/4 часов), их распределение по видам работ семестрам представлено в таблице 2.

ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость
	час. всего/*
Общая трудоёмкость дисциплины по учебному плану	144/4
1. Контактная работа	52,4/4
Аудиторная работа	52,4/4
лекции (Л)	16
практические занятия (ПЗ)	34/4
контактная работа на промежуточном контроле (КРА)	0,4
консультации	2
2. Самостоятельная работа (СРС)	64,6
самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям)	64,6
3. Контроль	27
Вид промежуточного контроля:	Экзамен

* в том числе практическая подготовка

4.2 Содержание дисциплины

Таблица 3

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ всего/*	ПКР	
Раздел 1 Введение в инжиниринг данных	26	4	6	-	16
Раздел 2 Процессы извлечения, преобразования и загрузки данных	30,6/2	4	10/2	-	16,6
Раздел 3 Хранилища данных и управление большими данными	30	4	10	-	16
Раздел 4 Оркестрация и качество данных	28/2	4	8/2	-	16
Консультации	2	-	-	2	-
Контроль	27	-	-	27	-
Контактная работа на промежуточном контроле	0,4	-	-	0,4	-
Итого по дисциплине	144	16	34/4	29,4	64,6

* в том числе практическая подготовка

Раздел 1 Введение в инжиниринг данных

Дает общее понимание роли инженера данных в экосистеме и фундаментальные навыки работы с базами данных и облачными платформами. Студенты осваивают основы SQL и NoSQL на практике. Рассматриваются такие темы, как: Введение в инжиниринг данных. Роль в Data-экосистеме; Основы работы с данными и СУБД.

Раздел 2 Процессы извлечения, преобразования и загрузки данных

Охватывает проектирование и построение пайплайнов для извлечения, преобразования и загрузки данных из различных источников. Рассматриваются такие темы, как: Проектирование и построение ETL/ELT процессов; Извлечение данных из различных источников. На практике отрабатывается создание ETL-скриптов и работа с инкрементальной загрузкой.

Раздел 3 Хранилища данных и управление большими данными

Знакомит с архитектурой и проектированием систем для хранения и анализа больших данных: Data Warehouses и Data Lakes. Рассматриваются такие темы, как: Хранилища данных и Data Lakes; Введение в экосистему Hadoop и Apache Spark. Практическая часть посвящена работе с облачными DWH и обработке данных с помощью Apache Spark.

Раздел 4 Оркестрация и качество данных

Учит автоматизации, мониторингу и обеспечению надежности пайплайнов с помощью оркестратора Apache Airflow. Рассматриваются такие темы, как: Оркестрация пайплайнов данных; Качество данных, мониторинг и DevOps для данных. Студенты создают автоматизированные workflows и добавляют в них проверки качества данных.

4.3 Лекции/практические занятия

Таблица 4

Содержание лекций/практических занятий и контрольные мероприятия

№ п/п	Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
1.	Раздел 1 Введение в инжиниринг данных		ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		10
	Тема 1 Введение в инжиниринг данных	Лекция 1. Введение в инжиниринг данных. Роль в Data-экосистеме	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Лекция 2. Основы работы с данными и СУБД	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Практическое занятие 1. Установка и настройка рабочего окружения (Docker, Python, IDE). Знакомство с облачной консолью	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы	2
		Практическое занятие 2. Основы SQL для инженера данных	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы	2
		Практическое занятие 3. Работа с NoSQL	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный	защита практической работы	2

№ п/п	Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
			уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		
2.	Раздел 2 Процессы извлечения, преобразования и загрузки данных		ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		14/2
	Тема 2 Процессы извлечения, преобразования и загрузки данных	Лекция 3. Проектирование и построение ETL/ELT процессов	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Лекция 4. Извлечение данных из различных источников	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Практическое занятие 4. Написание скрипта на Python для извлечения данных из публичного API Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 5. Веб-скрапинг простой страницы с помощью BeautifulSoup Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 6. Создание простого ETL-пайплайна Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2/2
		Практическое занятие 7. Реализация инкрементальной загрузки данных Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1;	защита практической работы, кейс-семинар	2

№ п/п	Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
			ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		
		Практическое занятие 8. Знакомство с визуальным инструментом ETL Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
3.	Раздел 3 Хранилища данных и управление большими данными		ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		14
	Тема 3 Хранилища данных и управление большими данными	Лекция 5. Хранилища данных и Data Lakes	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Лекция 6. Введение в экосистему Hadoop и Apache Spark	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Практическое занятие 9. Проектирование схемы «Звезда» Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 10. Настройка и выполнение запросов в облачном DWH Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 11. Основы работы с HDFS и Spark	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2.	защита практической работы,	2

№ п/п	Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Кейс-задачи от Россельхозбанка	Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	кейс-семинар	
		Практическое занятие 12. Выполнение операций ETL с помощью Spark DataFrame Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 13. Чтение и запись данных в различных форматах с использованием Spark. Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
4.	Раздел 4 Оркестрация и качество данных		ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2		12/2
	Тема 4 Оркестрация и качество данных	Лекция 7. Оркестрация пайплайнов данных	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Лекция 8. Качество данных, мониторинг и DevOps для данных	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1		2
		Практическое занятие 14. Установка и настройка Apache Airflow локально или в облаке. Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2

№ п/п	Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Практическое занятие 15. Написание первого DAG в Airflow для оркестрации простого ETL-пайплайна. Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2
		Практическое занятие 16. Создание сложного DAG с зависимостями и обработкой ошибок. Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2/2
		Практическое занятие 17. Реализация проверок качества данных в пайплайне. Кейс-задачи от Россельхозбанка	ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).2; ПККрмии-14 (ML-2. Экспертный уровень).2	защита практической работы, кейс-семинар	2

Таблица 5

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	Название раздела, темы	Перечень рассматриваемых вопросов для самостоятельного изучения
Раздел 1 Введение в инжиниринг данных		
1.	Тема 1 Введение в инжиниринг данных	CAP-теорема и ее практическое значение для выбора баз данных. Углубленное изучение оконных функций в SQL для сложной аналитики. Основы контейнеризации (Docker) для развертывания изолированных сред (например, базы данных). Написание сложных SQL-запросов с использованием подзапросов и CTE (Common Table Expressions) (ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1)
Раздел 2 Процессы извлечения, преобразования и загрузки данных		
2.	Тема 2 Процессы извлечения, преобразования и загрузки данных	Паттерны Change Data Capture (CDC) для отслеживания изменений в данных в реальном времени (например, с помощью Debezium). Знакомство с облачными ETL-сервисами (например, AWS Glue, Google Dataflow). Написание продвинутого парсера на Python (с использованием Scrapy или Selenium) для сбора данных со сложных веб-сайтов. Методы оптимизации ETL-процессов для работы с большими объемами данных (ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1)
Раздел 3 Хранилища данных и управление большими данными		
3.	Тема 3 Хранилища данных и управление большими данными	Углубленное изучение архитектурных паттернов Data Vault и Data Mesh. Изучение основ фреймворка Apache Kafka для построения потоковых данных (Data Streaming). Основы работы с кластером Hadoop (HDFS, YARN) и выполнение задач с помощью MapReduce. Изучение форматов колоночных данных (Apache Parquet, ORC) и их влияние на производительность (ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1)
Раздел 4 Оркестрация и качество данных		
4.	Тема 4 Оркестрация и качество данных	Принципы DataOps и MLOps для автоматизации жизненного цикла данных и моделей. Изучение альтернативных Airflow оркестраторов (например, Prefect, Dagster). Настройка мониторинга и алертинга для пайплайнов данных (например, при сбоях или падении качества данных). Основы управления доступом и шифрования данных в облачных хранилищах (ПККрмии-10 (BD-1. Экспертный уровень).1; ПККрмии-11 (BD-2. Экспертный уровень).1)

5. Образовательные технологии

Таблица 6

Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия		Наименование используемых активных и интерактивных образовательных технологий (форм обучения)
1.	Все темы	ПЗ	Сравнение технологий, case-study, проектное обучение

6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины

6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

Кейс-задачи от Россельхозбанка «Инжиниринг данных»

Кейс 1: Создание ETL-пайплайна для данных с датчиков сельхозтехники

Цель. Стандартизировать данные и загрузить в DWH для анализа эффективности использования техники.

Задача. Разработать ETL-процесс для сбора данных телеметрии (топливо, пробег, урожайность) с сельскохозяйственной техники, арендуемой через лизинговые программы банка. Данные поступают в формате JSON от разных производителей.

Кейс 2: Проектирование Data Vault для кредитного конвейера

Цель. Реализовать историчность данных и ускорить формирование отчетности для регуляторов.

Задача. Спроектировать модель хранения данных по кредитным заявкам фермерских хозяйств, обеспечивающую отслеживание изменений (например, статуса заявки, изменений в предоставленных документах).

Кейс 3: Инкрементальная загрузка данных о состоянии посевов

Цель. Минимизировать объем передаваемых данных и обеспечить актуальность информации для риск-менеджмента.

Задача. Настроить ежедневную инкрементальную загрузку спутниковых данных (NDVI-индекс) по полям заемщиков для мониторинга состояния урожая.

Кейс 4: Оркестрация пайплайна анализа платежной дисциплины

Цель. Автоматизировать процесс мониторинга кредитных рисков.

Задача. Создать Airflow DAG, который автоматически формирует еженедельный отчет по просроченным платежам среди заемщиков АПК, объединяя данные из Core Banking System и внешних источников (например, погодные данные для анализа влияния засух на платежеспособность).

Кейс 5: Построение Data Lake для неструктурированных данных

Цель. Создать единую платформу для последующего анализа с использованием компьютерного зрения и NLP.

Задача. Организовать хранение и каталогизацию разнородных данных: сканы договоров, спутниковые снимки полей, отчеты агрономов.

Кейс 6: Разработка пайплайна качества данных для справочника сельхозкультур

Цель. Исключить ошибки в данных, используемых для расчета лимитов кредитования.

Задача. Внедрить автоматические проверки (data quality) для обновляемого справочника сельскохозяйственных культур: контроль дубликатов, проверка корректности вегетационных периодов, валидация регионов выращивания.

Кейс 7: Реализация потокового пайплайна для мониторинга цен на зерно

Цель. Оперативно оценивать стоимость залогового имущества фермеров.

Задача. Настроить реальный time сбор данных о ценах на зерно с биржевых площадок и агрегаторов новостей с использованием Apache Kafka.

Кейс 8: Оптимизация запросов в DWH для отчетности по госсубсидиям

Цель. Сократить время формирования отчетов с 2 часов до 15 минут.

Задача. Проанализировать и оптимизировать медленные SQL-запросы, формирующие отчеты для Минсельхоза по объемам выданных субсидированных кредитов.

Кейс 9: Создание API для предоставления данных о почвах

Цель. Создать новый data-продукт для клиентов банка.

Задача. Разработать REST API, которое предоставляет агрономам данные о типах почв и истории севооборотов на основе объединения данных Россельхозбанка и открытых данных Росреестра.

Кейс 10: Миграция пайплайнов прогнозирования урожайности в облако

Цель. Снизить затраты на инфраструктуру и увеличить масштабируемость вычислений.

Задача. Перенести существующие ML-пайплайны прогнозирования урожайности (на основе исторических данных и погоды) из on-premise Hadoop кластера в облако (Russian Cloud Solutions).

**Перечень вопросов, выносимых на итоговую аттестацию
(экзамен)**

1. В чем заключается основная роль инженера данных в современной data-экосистеме?
2. Опишите жизненный цикл данных (Data Lifecycle).
3. В чем ключевые отличия между базами данных типа OLTP и OLAP?
4. Назовите основные типы NoSQL баз данных и приведите примеры использования для каждого.
5. Объясните, что такое CAP-теорема и как она влияет на выбор базы данных.
6. В чем разница между процессами ETL и ELT? Когда целесообразно использовать каждый из подходов?
7. Что такое инкрементальная загрузка данных и какие методы ее реализации вы знаете?
8. Какие существуют паттерны загрузки данных в хранилище? Опишите полную замену, инкрементальную загрузку и CDC.

9. Для чего используется инструмент Apache NiFi в контексте ETL?
10. Какие основные этапы включает в себя процесс веб-скрапинга и с какими юридическими и этическими ограничениями можно столкнуться?
11. Опишите принцип работы и преимущества Change Data Capture (CDC).
12. Что такое Data Warehouse? Перечислите его ключевые характеристики.
13. Опишите различия между схемами «Звезда» и «Снежинка».
14. Что такое Data Lake и в чем его основное отличие от Data Warehouse?
15. В чем заключается концепция Data Lakehouse и какие проблемы она призвана решить?
16. Каковы основные компоненты архитектуры Hadoop (HDFS, YARN, MapReduce)?
17. Что такое Apache Spark и какие преимущества он дает по сравнению с Hadoop MapReduce?
18. Опишите основные абстракции данных в Spark: RDD, DataFrame, Dataset.
19. Какие форматы хранения данных вы знаете (Parquet, Avro, ORC) и в чем их преимущества перед CSV/JSON?
20. Для чего предназначен Apache Airflow и что такое DAG?
21. Опишите ключевые концепции Airflow: Operator, Task, Dependency.
22. Как в Airflow обеспечивается оркестрация сложных пайплайнов с зависимостями?
23. Что такое Data Quality (качество данных) и какие метрики для его оценки вы знаете?
24. Какие инструменты для тестирования и мониторинга качества данных вы можете назвать?
25. Объясните, что такое DataOps и как его принципы применяются в инжиниринге данных.
26. Как обеспечивается надежность (reliability) и отказоустойчивость пайплайнов данных?
27. Какие стратегии мониторинга пайплайнов данных вы знаете?
28. Как вы выбираете формат хранения данных для конкретной задачи?
29. Опишите процесс проектирования и реализации пайплайна для сбора данных из API.
30. Какие best practices по обеспечению безопасности данных в облачной среде вы знаете?
31. Что такое потоковая обработка данных (stream processing) и чем она отличается от пакетной (batch processing)?
32. Для чего используется Apache Kafka и каково ее место в data-пайплайне?
33. Что такое контейнеризация и как Docker используется в инжиниринге данных?
34. Как вы организуете версионирование данных и кода пайплайнов?
35. Опишите процесс миграции пайплайна данных из тестовой среды в продакшен (production).
36. Какие вы знаете облачные сервисы для построения ETL-процессов?
37. Что такое «feature store» и какую проблему он решает в машинном обучении?
38. Как вы подходите к оптимизации медленно работающего SQL-запроса?

39. Какие методы оптимизации производительности Spark приложений вы знаете?
40. Как работает механизм партиционирования данных в Spark и HDFS и зачем он нужен?
41. Что такое шифрование данных at-rest и in-transit?
42. Как вы управляете доступом к данным в рамках большой команды?
43. Опишите процесс автоматического перезапуска упавшего пайплайна в Airflow.
44. Что такое Data Lineage (происхождение данных) и почему это важно?
45. Каковы основные этапы планирования проекта по инжинирингу данных?
46. С какими основными вызовами и проблемами можно столкнуться при работе с большими данными (Big Data)?
47. Как вы обеспечиваете воспроизводимость (reproducibility) результатов работы пайплайна?
48. Опишите, как бы вы спроектировали систему для сбора и анализа логов веб-приложения.
49. В чем разница между инструментами оркестрации Airflow и Prefect?
50. Какие тренды в области инжиниринга данных вы считаете наиболее перспективными на ближайшие годы?

Пример работ

Практическая работа № 1. Установка и настройка рабочего окружения (Docker, Python, IDE). Знакомство с облачной консолью

Цель работы. Сформировать базовое рабочее окружение инженера данных, освоить основные инструменты контейнеризации и облачной инфраструктуры, получить практический опыт развертывания изолированной среды для работы с данными.

Краткое содержание работы. Необходимо последовательно выполнить установку и настройку ключевых инструментов. Сначала требуется установить Docker, создав контейнер с базой данных PostgreSQL, проверив его работу через подключение и выполнение базовых SQL-запросов. Далее нужно установить интерпретатор Python версии 3.8 или выше, настроить виртуальное окружение и установить в него основные библиотеки для работы с данными: pandas для анализа, sqlalchemy для взаимодействия с базами данных, jupyter для интерактивной разработки. После этого следует установить и настроить интегрированную среду разработки PyCharm Community Edition или VS Code, подключив их к созданному виртуальному окружению и настроив плагины для работы с Docker и Jupyter. Финальный этап – регистрация в облачной платформе AWS по программе Free Tier, изучение интерфейса консоли управления, создание виртуальной машины EC2 с базовыми параметрами и объекта в сервисе хранения S3 с загрузкой тестового файла.

Итоговый результат. Полностью настроенное локальное окружение с работающим Docker-контейнером PostgreSQL, настроенным Python-окружением с

ключевыми библиотеками, интегрированной средой разработки и созданной базовой облачной инфраструктурой в AWS. Студент должен продемонстрировать работоспособность всех компонентов: подключение к контейнеру с БД, выполнение Python-скрипта с импортом библиотек, запуск Jupyter-ноутбука и доступ к созданным облачным ресурсам в AWS-консоли.

Практическая работа № 2. Основы SQL для инженера данных

Цель работы. Сформировать навыки создания базы данных, написания базовых SQL-запросов и использования операций соединения таблиц и агрегирующих функций для анализа данных.

Краткое содержание работы. Создать базу данных продаж, состоящую из двух таблиц - товаров и заказов. Для таблицы products необходимо создать колонки: product_id (целое число, первичный ключ), product_name (текст), category (текст), price (дробное число). Для таблицы orders создать колонки: order_id (целое число, первичный ключ), product_id (целое число, внешний ключ), quantity (целое число), order_date (дата).

Заполните таблицы тестовыми данными: 5-7 товаров разных категорий (например, «Электроника», «Одежда», «Книги») и 10-15 записей о заказах с разными датами и количествами товаров.

Написать серию SQL-запросов: выборка всех товаров определенной категории; подсчет общего количества проданных единиц для каждого товара с использованием JOIN и GROUP BY; поиск товаров, проданных в количестве более 5 штук за один заказ; расчет общей выручки по категориям товаров; выборка заказов, содержащих товары дороже 500 рублей через подзапрос или JOIN.

Итоговый результат. Готовый SQL-файл, содержащий последовательность команд создания базы данных, создания таблиц, наполнения их данными и все требуемые запросы с комментариями. Каждый запрос должен быть проверен на корректность выполнения и возвращать ожидаемый результат на тестовых данных. Для проверки работы запросов необходимо предоставить скриншоты выполнения или текстовые выдачи результатов.

Практическая работа № 3. Работа с NoSQL

Цель работы. Освоить базовые операции CRUD (Create, Read, Update, Delete) в системе управления базами данных NoSQL на примере MongoDB или Redis.

Краткое содержание работы. Установить и настроить выбранную СУБД NoSQL (MongoDB или Redis). Для MongoDB: создать базу данных shop_db и коллекцию products. Выполнить операции: добавить 5-7 документов с информацией о товарах (поля: name, category, price, quantity), найти все товары определенной категории, обновить цену для конкретного товара, увеличить количество на складе для всех товаров категории, удалить товары с нулевым остатком. Для Redis: работать с хешами для хранения информации о товарах

(поля: name, category, price, quantity). Выполнить: добавление 3-5 товаров, чтение всех полей товара, обновление цены, удаление одного товара, получение всех ключей.

Итоговый результат. Скрипт или файл с последовательностью команд CRUD для выбранной СУБД NoSQL с комментариями. Проверка работоспособности каждой операции через вывод результатов выполнения. Для MongoDB - примеры документов до и после изменений, для Redis - скриншоты работы в клиенте redis-cli или другом интерфейсе.

6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания

Для оценки знаний, умений, навыков и формирования компетенции по дисциплине применяется **балльно-рейтинговая** система контроля и оценки успеваемости студентов.

В основу балльно-рейтинговой системы (БРС) положены принципы, в соответствии с которыми формирование рейтинга студента осуществляется в ходе текущей работы в семестре.

По результатам защиты могут быть получены следующие баллы.

9-10 баллов – расчеты (если имеются) проведены корректно, результаты правильно интерпретированы. Полностью выполнены все пункты выданного задания. Работа оформлена в соответствии с требованиями стандартов по оформлению текстовых документов. Студент развернуто и свободно ответил на все вопросы при защите работы.

7-8 баллов – работа выполнена, выполнены все пункты выданного задания, но не полностью, либо с несущественными ошибками, имеются незначительные ошибки в интерпретации результатов и/или оформлении. Студент в целом ответил на все поставленные вопросы, ориентируется в работе.

4-6 баллов – работа в целом выполнена, выполнены основные, но не все пункты выданного задания, либо с существенными ошибками, имеются значительные ошибки в интерпретации полученных результатов и представления данных, оформления работы. Некоторые вопросы по работе вызывают затруднения.

1-3 балла – имеются грубые ошибки в методике выполнения, интерпретации полученных результатов и представления данных, оформления работы, большая часть пунктов выданного задания не выполнена. Студент не отвечает на вопросы при защите.

Итоговая оценка по дисциплине выставляется преподавателем в соответствии со шкалой.

Таблица 7

Текущий рейтинг	Оценка			
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
в процентах	0-59	60-69	70-84	85-100

К сдаче экзамена допускаются студенты сдавшие все практические работы.

Критерии оценивания результатов обучения

Таблица 8

Оценка	Критерии оценивания
Высокий уровень «5» (отлично)	оценку «отлично» заслуживает студент, освоивший знания, умения, компетенции и теоретический материал без пробелов; выполнивший все задания, предусмотренные учебным планом на высоком качественном уровне; практические навыки профессионального применения освоенных знаний сформированы. Компетенции, закреплённые за дисциплиной, сформированы на уровне – высокий.
Средний уровень «4» (хорошо)	оценку «хорошо» заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформировал практические навыки. Компетенции, закреплённые за дисциплиной, сформированы на уровне – хороший (средний).
Пороговый уровень «3» (удовлетворительно)	оценку «удовлетворительно» заслуживает студент, частично с пробелами освоивший знания, умения, компетенции и теоретический материал, многие учебные задания либо не выполнил, либо они оценены числом баллов близким к минимальному, некоторые практические навыки не сформированы. Компетенции, закреплённые за дисциплиной, сформированы на уровне – достаточный.
Минимальный уровень «2» (неудовлетворительно)	оценку «неудовлетворительно» заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы. Компетенции, закреплённые за дисциплиной, не сформированы.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1 Основная литература

1. Советов, Б. Я. Информационные технологии : учебник для вузов / Б. Я. Советов, В. В. Цехановский. — 8-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 414 с. — (Высшее образование). — ISBN 978-5-534-20054-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559897>
2. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2025. — 97 с. — (Высшее образование). — ISBN 978-5-534-21173-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559502>
3. Советов, Б. Я. Базы данных : учебник для вузов / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 403 с. — (Высшее образование). — ISBN 978-5-534-18479-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559898>.

4. Мамедли, Р. Э. Большие данные и NoSQL базы данных : учебное пособие для вузов / Р. Э. Мамедли, Т. Б. Казиахмедов. — Санкт-Петербург : Лань, 2024. — 92 с. — ISBN 978-5-507-49873-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/434051>

7.2 Дополнительная литература

1. Чак, Л. Hadoop в действии / Л. Чак. — Москва : ДМК Пресс, 2012. — 424 с. — ISBN 978-5-94074-785-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/39997>

2. Перрен, Ж. -. Spark в действии / Ж. -. Перрен ; перевод с английского А. В. Снастина. — Москва : ДМК Пресс, 2021. — 636 с. — ISBN 978-5-97060-879-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/241001>

3. Харенслак, Б. Apache Airflow и конвейеры обработки данных / Б. Харенслак, Р. Д. де ; перевод с английского Д. А. Беликова. — Москва : ДМК Пресс, 2022. — 502 с. — ISBN 978-5-97060-970-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/241133>

4. Макшанов, А. В. Системы поддержки принятия решений : учебное пособие для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 108 с. — ISBN 978-5-8114-8489-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176903>

7.3 Статьи, опубликованные в научных журналах 1 уровня Белого списка научных журналов Минобрнауки России и сборниках научных работ конференций уровня А*

1. Petroni, F., Broscheit, S., Piktus, A. et al. Improving Wikipedia verifiability with AI. Nat Mach Intell 5, 1142–1148 (2023). <https://doi.org/10.1038/s42256-023-00726-1>

2. Материалы конференции Transactions of the association for computational linguistics 2024. — URL: <https://aclanthology.org/volumes/2024.tacl-1/>

3. Материалы конференции Very Large Data Base Endowment Inc. (VLDB Endowment). — URL: <https://vldb.org/>

4. Материалы конференции ACM SIGMOD International Conference on Management of Data. — URL: <https://sigmod.org/>

5. Материалы конференции IEEE International Conference on Data Engineering. — URL: <https://waset.org/data-engineering-conference>

6. Материалы конференции Conference on Innovative Data Systems Research. — URL: <https://www.cidrdb.org/>

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Apache Software Foundation. — URL: <https://apache.org/> (Документация по Hadoop, Spark, Airflow, Kafka)

2. GitHub. – URL: <https://github.com/> (Open-source инструменты для инжиниринга данных)
3. Kaggle. – URL: <https://kaggle.com/> (Датасеты для практики)
4. Google Cloud. – URL: <https://cloud.google.com> (Документация по облачным DWH)
5. Microsoft Azure. – URL: <https://azure.microsoft.com/ru-ru/> (Демодоступ к Azure Data Factory)
6. Habr. – URL: <https://habr.com/ru> (Статьи и кейсы по инжинирингу данных)
7. Docker Hub. – URL: <https://hub.docker.com> (Контейнеры с предустановленными инструментами)
8. Apache Airflow GitHub. – URL: <https://github.com/apache/airflow> (Примеры DAG и лучшие практики)

9. Перечень программного обеспечения и информационных справочных систем

Таблица 9

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Разделы 1, 2, 3,4	Python	расчетная, обучающая, контролирующая	Python Software Foundation	Текущая версия
		Docker	контейнеризация	Docker, Inc.	Текущая версия
		PostgreSQL	СУБД	PostgreSQL Global Development Group	Текущая версия
		MongoDB	NoSQL СУБД	MongoDB, Inc.	Текущая версия
		Apache Hadoop	Фреймворк распределенной обработки данных	Apache Software Foundation	Текущая версия
		Apache Spark	Фреймворк для распределенной обработки больших данных		
		Apache Airflow	Платформа оркестрации workflows		
		DBeaver	Клиент для баз данных	DBeaver Community	Текущая версия
		SQLite	СУБД	SQLite Consortium	Текущая версия

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Занятия проводятся в специализированных аудиториях, оснащенных мультимедийным оборудованием для проведения лекционных занятий. Практические занятия проводятся с использованием технических и программных средств в аудитории, оснащенной персональными компьютерами и доступом в интернет.

Таблица 10

Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями

Наименование специальных помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы
1	2
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (2й учебный корпус, 102 ауд.)</i>	Количество рабочих мест. 16 1. Компьютеры 28 шт. Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа. Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. 2. Стенд «Сергеев Сергей Степанович 1910-1999» 1 шт. (Инв. №591013/25) 3. Огнетушитель порошковый 1 шт. (Инв. №559527) 4. Подвесное крепление к огнетушителю 1 шт. (Инв. № 559528) 5. Жалюзи 2шт. (Инв. №1107-221225, Инв. №1107-221225) 6. Стул 29 шт. 7. Стол компьютерный 28 шт. 8. Стол для преподавателя 1 шт. 9. Доска маркерная 1 шт. 10. Трибуна напольная 1 шт. (без инв. №) Структурное подразделение. Институт Экономики и управления, Кафедра Статистики и кибернетики
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория</i>	Количество рабочих мест. 16 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа. Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение. Институт Экономики и управления, Кафедра Статистики и кибернетики

для текущего контроля и промежуточной аттестации (2й учебный корпус, 106 ауд.)	
учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы (2й учебный корпус, 302 ауд.)	Количество рабочих мест. 16 1. Системный блок 17 шт. - Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа. Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. 2. Монитор 17 шт. 3. Телевизор 1 шт. 4. Стол для преподавателя 1 шт. 5. Стол компьютерный 16 шт. 6. Стул офисный 17 шт. Структурное подразделение. Институт Экономики и управления, Кафедра Статистики и кибернетики
учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (1й учебный корпус, 212 ауд.)	Количество рабочих мест. 24 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа. Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение. Кафедра Цифровая кафедра
учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (1й учебный корпус, 214 ауд.)	Количество рабочих мест. 24 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа. Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение. Кафедра Цифровая кафедра
Центральная научная библиотека имени Н.И. Железнова	Читальные залы библиотеки
Студенческое общежитие	Комната для самоподготовки

11. Методические рекомендации обучающимся по освоению дисциплины

Предполагается, что студент выполняет практическое задание в аудитории, дома оформляет и готовится по теоретическим вопросам к защите отчета на следующем занятии.

Виды и формы отработки пропущенных занятий

Студент, пропустивший занятие, обязан предъявить преподавателю документы установленного образца, подтверждающие необходимость пропуска. Не допускается пропуск занятий без уважительной причины.

Студент, пропустивший занятия, осваивает материал самостоятельно (выполняет практическое задание по своему варианту в компьютерном классе кафедры в часы, свободные от занятий, изучает теоретические вопросы).

Студент, пропустивший лекцию, отвечает на вопросы по пропущенной теме.

12. Методические рекомендации преподавателям по организации обучения по дисциплине

На первом занятии преподаватель закрепляет за каждым студентом номер варианта для выполнения индивидуальных работ (как правило, номер варианта соответствует порядковому номеру студента в журнале преподавателя). По каждой индивидуальной работе должна быть поставлена оценка по факту ее защиты. Защиту рекомендуется проводить на следующем после получения задания занятии. Преподаватель обязан проверить соответствие выполненного задания исходным данным варианта студента. Таким образом, исключается вероятность плагиата.

Преподаватель должен стимулировать студентов к занятию научно-исследовательской работой, изучению научной литературы по теме искусственного интеллекта, в т.ч. отечественной и зарубежной периодики.

Программу разработал:

Уколова А.В., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)

Бодур А.М., ассистент
(ФИО, ученая степень, ученое звание)



(подпись)

(подпись)

РЕЦЕНЗИЯ
на рабочую программу дисциплины «Инжиниринг данных»
ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии»,
направленность «Большие данные и машинное обучение»
(квалификация выпускника – бакалавр)

Вахрушевой Инной Алексеевной, доцентом кафедры высшей математики ФГБОУ ВО «Российский государственный аграрный университет - МСХА имени К.А. Тимирязева, кандидатом педагогических наук (далее по тексту рецензент), проведена рецензия рабочей программы дисциплины «Методы машинного обучения» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Большие данные и машинное обучение», разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчики – Уколова А.В., к.э.н., доцент кафедры статистики и кибернетики, Бодур А.М., ассистент кафедры статистики и кибернетики)

Рассмотрев представленные на рецензию материалы, рецензент пришел к следующим выводам.

1. Предъявленная рабочая программа дисциплины «Инжиниринг данных» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.03.02 «Информационные системы и технологии» и компетентностно-ролевых моделей в сфере искусственного интеллекта (КРМ ИИ), разработанной по заказу Аналитического центра при Правительстве Российской Федерации в рамках федерального проекта «Искусственный интеллект» национального проекта «Экономика данных и цифровая трансформация государства». Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

2. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к части учебного цикла, формируемой участниками образовательных отношений – Б1.В.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии» и компетентностно-ролевых моделей в сфере искусственного интеллекта (КРМ ИИ).

4. В соответствии с учебным планом и компетентностно-ролевыми моделями в сфере искусственного интеллекта (КРМ ИИ) за дисциплиной «Инжиниринг данных» закреплена 3 инструментальных компетенций (4 индикатора). Дисциплина «Инжиниринг данных» и представленная Программа способна реализовать их в объявленных требованиях. Результаты обучения, представленные в программе, соответствуют специфике и содержанию дисциплины и демонстрируют возможность получения заявленных результатов.

5. Общая трудоёмкость дисциплины «Инжиниринг данных» составляет 4 зачётные единицы (144 часа, в т.ч. 4 часа практической подготовки).

6. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Инжиниринг данных» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии» и возможность дублирования в содержании отсутствует.

7. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

8. Программа дисциплины «Инжиниринг данных» предполагает проведение занятий в интерактивной форме.

9. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

10. Представленные и описанные в Программе формы *текущей* оценки знаний (защита практических работ), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета в первом семестре и экзамена во втором семестре изучения, что соответствует статусу дисциплины, как дисциплины части учебного цикла, формируемой участниками образовательных отношений – Б1.В ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

11. Формы оценки знаний, представленные в Программе, соответствуют специфике дисциплины и требованиям к выпускникам.

12. Учебно-методическое обеспечение дисциплины представлено. основной литературой – 4 источника (базовый учебник), дополнительной литературой – 4 наименований, статьи, опубликованные в научных журналах 1 уровня Белого списка научных журналов Минобрнауки России и сборниках научных работ конференций уровня А* - 6 источников, Интернет-ресурсы – 8 источников и соответствует требованиям ФГОС ВО и компетентностно-ролевых моделей в сфере искусственного интеллекта направления 09.03.02 «Информационные системы и технологии».

13. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Инжиниринг данных» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

14. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Инжиниринг данных».

ОБЩИЕ ВЫВОДЫ

На основании проведенной рецензии можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Инжиниринг данных» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Большие данные и машинное обучение» (квалификация выпускника – бакалавр), разработанная Уколовой А.В., к.э.н., доцентом кафедры статистики и кибернетики, Бодур А.М., ассистентом кафедры статистики и кибернетики, соответствует требованиям ФГОС ВО и компетентностно-ролевых моделей в сфере искусственного интеллекта (КРМ ИИ), разработанной по заказу Аналитического центра при Правительстве Российской Федерации в рамках федерального проекта «Искусственный интеллект» национального проекта «Экономика данных и цифровая трансформация государства», современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент: Вахрушева Инна Алексеевна, доцент кафедры высшей математики ФГБОУ ВО «Российский государственный аграрный университет - МСХА имени К.А. Тимирязева», кандидат педагогических наук


(подпись)