

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Структурное подразделение «Брикс кафедры (205)»

УТВЕРЖДЕНА:
на заседании кафедры
Протокол №15 от 18 марта 2025 г.

Рабочая программа дисциплины

«БОЛЬШИЕ ДАННЫЕ, МЕТОДЫ ОБРАБОТКИ / BIG DATA, PROCESSING METHODS»

Направление: 09.03.01 Информатика и вычислительная техника

Искусственный интеллект и компьютерные науки /Artificial Intelligence and Computer
Science

Квалификация: Бакалавр

Форма обучения: очная

Документ подписан простой
электронной подписью
Составитель программы:
Григорьев Станислав
Валентинович
Дата подписания: 09.06.2026

Документ подписан простой
электронной подписью
Утвердил: Киреев Анна
Павловна
Дата подписания: 09.06.2026

Документ подписан простой
электронной подписью
Согласовал: Афанасьев
Александр Диомидович
Дата подписания: 09.06.2026

Год набора – 2026

Иркутск, 2025 г.

1 Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы

1.1 Дисциплина «Большие данные, методы обработки / Big Data, Processing Methods» обеспечивает формирование следующих компетенций с учётом индикаторов их достижения

Код, наименование компетенции	Код индикатора компетенции
ПКС-1 Способен проводить самостоятельные исследования в области искусственного интеллекта	ПКС-1.1
ПКС-5 Способен реализовывать проекты в сфере искусственного интеллекта	ПКС-5.1

1.2 В результате освоения дисциплины у обучающихся должны быть сформированы

Код индикатора	Содержание индикатора	Результат обучения
ПКС-1.1	Осуществляет сбор и анализ теоретического материала для решения исследовательской задачи	Знать Знать стандарты обработки и анализа больших данных, и требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных. Уметь Уметь визуализировать, интерпретировать и давать рекомендации на основании результатов исследования и предварительной обработки больших данных. Владеть Владеть способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.
ПКС-5.1	Понимает методы обработки больших данных, их применимость и ограничения при реализации проектов в сфере профессиональной деятельности	Знать Знать методику проведения исследований и подготовки неструктурированных и полуструктурированных данных для анализа. Уметь Уметь осуществлять исследования неструктурированных и полуструктурированных данных и анализировать их результаты. Владеть Владеть навыками проведения исследований и обработки неструктурированных и полуструктурированных данных.

2 Место дисциплины в структуре ООП

Изучение дисциплины «Большие данные, методы обработки / Big Data, Processing Methods» базируется на результатах освоения следующих дисциплин/практик: «Математическая логика и теория алгоритмов / Mathematical Logic and Algorithm Theory», «Вычислительная математика / Computational Mathematics», «Программирование / Programming», «Математика / Mathematics»

Дисциплина является предшествующей для дисциплин/практик: «Технологии внедрения решений искусственного интеллекта / Artificial Intelligence Implementation Technologies», «Технологии разработки программного обеспечения / Software Development Technologies», «Интеллектуальный анализ данных / Data Mining», «Распределённые вычисления и блокчейн-приложения / Distributed Computing and Blockchain Applications», «Теоретические основы искусственного интеллекта / Theoretical Foundations of Artificial Intelligence»

3 Объем дисциплины

Объем дисциплины составляет – 5 ЗЕТ

Вид учебной работы	Трудоемкость в академических часах (Один академический час соответствует 45 минутам астрономического часа)	
	Всего	Семестр № 3
Общая трудоемкость дисциплины	180	180
Аудиторные занятия, в том числе:	60	60
лекции	30	30
лабораторные работы	0	0
практические/семинарские занятия	30	30
Самостоятельная работа (в т.ч. курсовое проектирование)	84	84
Трудоемкость промежуточной аттестации	36	36
Вид промежуточной аттестации (итогового контроля по дисциплине)	Экзамен	Экзамен

4 Структура и содержание дисциплины

4.1 Сводные данные по содержанию дисциплины

Семестр № 3

№ п/п	Наименование раздела и темы дисциплины	Виды контактной работы						СРС		Форма текущего контроля
		Лекции		ЛР		ПЗ(СЕМ)				
		№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	
1	2	3	4	5	6	7	8	9	10	11
1	Ведение в большие данные	1	4			1	2	1, 2, 3	12	Тест
2	Сбор и хранение больших данных	2	4			2	4	1, 2, 3, 4	16	Тест
3	Методы обработки и анализа больших	3	6			3, 5, 6	12	1, 2, 3, 4	16	Тест

	данных									
4	Современные программные средства анализа больших объемов информации	4	6			7	4	1, 2, 3	12	Тест
5	Облачные вычисления для обработки больших данных	5	6			8	4	1, 2, 3, 4	16	Тест
6	Визуализация исходной информации и аналитических данных	6	4			4	4	1, 2, 3	12	Тест
	Промежуточная аттестация								36	Экзамен
	Всего		30				30		120	

4.2 Краткое содержание разделов и тем занятий

Семестр № 3

№	Тема	Краткое содержание
1	Ведение в большие данные	Введение в большие данные. Обзор источников информации. Описание концепции и архитектуры Обработка больших данных. Решение для обработка больших данных.
2	Сбор и хранение больших данных	Технологии хранения и обработки больших. Сложности построения хранилищ данных. Архитектура хранилищ данных.
3	Методы обработки и анализа больших данных	Статистические методы анализа данных. Методы многомерной оптимизации для построения оценок. Статистические и систематические ошибки. Использование языка Python для анализа данных. Байесовские методы. Методы Монте-Карло для анализа данных
4	Современные программные средства анализа больших объемов информации	Введение в ETL с SSIS. Изучение источников данных. Реализация потока данных. Введение в поток управления. Создание динамических пакетов. Использование контейнеров. Управление согласованностью. Введение в инкрементные процессы ETL. Извлечение изменения данных. Загрузка измененных данных.
5	Облачные вычисления для обработки больших данных	Обзор «облачных» источников данных SQL Server Azure. Использование PowerPivot для создания запроса к хранилищу данных, визуализация данных с помощью Основные понятия Master Data Services (MDS), реализация модели Master Data Services, использование надстройки Excel с Master Data Services.
6	Визуализация исходной информации и аналитических данных	Способы визуального представления данных. Методы визуализации. Представление пространственных характеристик. Разработка

		сложных видов диаграмм. Повышение уровня взаимодействия с визуализацией пользователя. Увеличение размеров и сложности структур данных, представляемых визуализацией.
--	--	--

4.3 Перечень лабораторных работ

Лабораторных работ не предусмотрено

4.4 Перечень практических занятий

Семестр № 3

№	Темы практических (семинарских) занятий	Кол-во академических часов
1	Ознакомление с синтаксисом языка R для анализа данных	2
2	Способы подготовки и отображения данных в R. Возможности ввода/вывода	4
3	Способы анализа данных в R. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных	4
4	Работа с диаграммами и графиками в R	4
5	Корреляционный анализ и регрессионный анализ данных	4
6	Задачи классификации, кластеризации: деревья решений, RandomForest, k-means	4
7	Изучение принципов работы распределенных баз данных	4
8	Развертывание локального кластера Hadoop. Подсчёт слов в тексте, с помощью MapReduce	4

4.5 Самостоятельная работа

Семестр № 3

№	Вид СРС	Кол-во академических часов
1	Оформление отчетов по лабораторным и практическим работам	24
2	Подготовка к практическим занятиям	24
3	Подготовка к сдаче и защите отчетов	24
4	Проработка разделов теоретического материала	12

В ходе проведения занятий по дисциплине используются следующие интерактивные методы обучения: В ходе проведения лекций, практических и лабораторных работ используются следующие интерактивные методы обучения интерактивная демонстрация способа решения типовой проблемы, метод проектов, лекция-дискуссия.

5 Перечень учебно-методического обеспечения дисциплины

5.1 Методические указания для обучающихся по освоению дисциплины

5.1.1 Методические указания для обучающихся по практическим занятиям

Григорьев С.В. Big Data Processing Methods / Электронный курс/ URL:
<https://el.istu.edu/course/view.php?id=7325>

5.1.2 Методические указания для обучающихся по самостоятельной работе:

1. Lectures on scientific computing with Python. В свободном доступе: URL:
<https://github.com/jrjohansson/scientific-python-lectures>
2. Мерков, Александр Борисович. Распознавание образов: введение в методы статистического обучения / А.Б. Мерков; Рос. акад. наук, Ин-т систем. анализа. — Москва: УРСС=URSS, 2010.— 254с.: ил. (URL:
<http://www.recognition.mccme.ru/pub/RecognitionLab.html/slbook.pdf>
3. Доклад ЦПУ про большие задачи и большие данные: URL: <http://bit.ly/CRUbigdata>
4. Тренажер для освоения основ языка R: <http://tryr.codeschool.com/> от O'Reilly
5. Язык R: из учебной лаборатории — в мир больших данных. URL: Леонид Черняк.
osp.ru/os/2012/04/13015768/

6 Фонд оценочных средств для контроля текущей успеваемости и проведения промежуточной аттестации по дисциплине

6.1 Оценочные средства для проведения текущего контроля

6.1.1 семестр 3 | Тест

Описание процедуры.

Последовательность тестовых случаев в порядке выполнения и любые связанные с ними действия, которые могут потребоваться для настройки начальных предварительных условий, а также любые завершающие действия после выполнения.

Критерии оценивания.

Критерии оценки теста включают в себя количество правильных ответов и соответствие заданным требованиям. Оценка "отлично" ставится за 90-100% правильных ответов, "хорошо" за 70-89%, "удовлетворительно" за 50-69%, и "неудовлетворительно" за менее 50%.

6.2 Оценочные средства для проведения промежуточной аттестации

6.2.1 Критерии и средства (методы) оценивания индикаторов достижения компетенции в рамках промежуточной аттестации

Индикатор достижения компетенции	Критерии оценивания	Средства (методы) оценивания промежуточной аттестации
ПКС-1.1	Демонстрирует умение управлять ИТ-инфраструктурой, информационными процессами и системами.	Отчеты по лабораторным работам, вопросы для контроля по

		лабораторным работам.
ПКС-5.1	Обнаруживает всестороннее, систематическое и глубокое знание учебно-программного материала; принимает активное участие в обсуждении по теме занятия; усвоил основную и дополнительную литературу, рекомендованную рабочей программой дисциплины; проявляет творческие способности в понимании, изложении и использовании учебно-программного материала.	Отчеты по лабораторным работам, вопросы для контроля по лабораторным работам.

6.2.2 Типовые оценочные средства промежуточной аттестации

6.2.2.1 Семестр 3, Типовые оценочные средства для проведения экзамена по дисциплине

6.2.2.1.1 Описание процедуры

Экзамен проводится в устной и письменной форме по билетам, составленных из вопросов.

1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.
2. Дать краткую сравнительную характеристику инструментария ПО для анализа данных.
3. Охарактеризовать конструкции языка R перечислить типы языка R, привести примеры.
4. Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика. Отличия BI от Data Science.
5. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.
6. Что такое Data Mining? Основные задачи и методы Data Mining. Этапы интеллектуального анализа данных. Методы интеллектуального анализа данных.
7. Статические и динамические знания.
8. Роль гипотез в процессе познания. Какие факторы используются для уточнения гипотез?
9. Основные понятия статистики и дескриптивный анализ:
10. Шкалы измерений. Генеральная совокупность и выборка. Нормальное распределение. Уровень статистической достоверности.
11. Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа.
12. Пояснить термин "Линейная регрессия". Привести примеры использования регрессионного анализа.
13. Классификация и кластеризация – суть и назначение. Метрики. Постановка задачи кластеризации. Методы кластеризации на графах. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.
14. Парадигма Map Reduce. Описать принцип работы. Нарисовать схему. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.

15. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры и инструменты для визуализации.
16. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.
17. OLAP и OLTP системы. Разница.
18. Репликация и шардинг.
19. Требования ACID. CAP-теорема, BASE архитектура
20. NoSQL. Классификация NoSQL хранилищ. Их особенности. Примеры распределенных хранилищ.

Пример задания:

Билет 1.

1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных.
2. Дать краткую сравнительную характеристику инструментария ПО для анализа данных.
3. Охарактеризовать конструкции языка R перечислить типы языка R, привести примеры.

6.2.2.1.2 Критерии оценивания

Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
Студент умеет увязывать теорию с практикой (решает задачи и формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, обосновывает свои суждения и даёт правильные ответы на вопросы преподавателя	Студент умеет увязывать теорию с практикой (решает задачи и формулирует выводы, умеет пояснить полученные результаты), владеет понятийным аппаратом, полно и глубоко овладел материалом по заданной теме, но содержание ответов имеют некоторые неточности и требуют уточнения и комментария со стороны преподавателя	Студент знает и понимает материал по заданной теме, но изложение неполное, непоследовательное, допускаются неточности в определении понятий, Студент не может обосновать свои ответы на уточняющие вопросы преподавателя	Студент допускает ошибки в определении понятий, искажающие их смысл, беспорядочно и неуверенно излагает материал. Делает ошибки в ответах на уточняющие вопросы преподавателя.

7 Основная учебная литература

1. Big Data Governance and Perspectives in Knowledge Management / edited by Sh. K. Strydom, M. Strydom, 2019. - 320.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37062.pdf>

2. Marin I. Big Data Analysis with Python. Combine Spark and Python to Unlock the Powers of Parallel Computing and Machine Learning / I. Marin, A. Shukla, S. Vk, 2019. - 276.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37559.pdf>

3. Handbook of Research on Big Data Clustering and Machine Learning / ed. F. P. Garcia Marquez, 2020. - 500.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-38071.pdf>

4. Handbook of Research on Lifestyle Sustainability and Management Solutions Using AI, Big Data Analytics, and Visualization / ed.: S. S. Iyer [et al.], 2021. - 438.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-40105.pdf>

8 Дополнительная учебная литература и справочная

1. Deep Learning Techniques and Optimization Strategies in Big Data Analytics / ed.: J. J. Thomas [et al.], 2020. - 381.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-38064.pdf>

2. Sedkaoui S. Big Data Analytics for Entrepreneurial Success / S. Sedkaoui, 2019. - 322.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37053.pdf>

3. Emerging Perspectives in Big Data Warehousing / edited by D. Taniar, J. W. Rahayu, 2019. - 367.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37897.pdf>

4. Big Data Analytics for Improved Accuracy, Efficiency, and Decision Making in Digital Marketing / ed. A. Singh, 2021. - 335.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-39492.pdf>

9 Ресурсы сети Интернет

1. <http://library.istu.edu/>
2. <https://e.lanbook.com/>

10 Профессиональные базы данных

1. <http://new.fips.ru/>
2. <http://www1.fips.ru/>

11 Перечень информационных технологий, лицензионных и свободно распространяемых специализированных программных средств, информационных справочных систем

1. Windows Server 2019 (Standard Core/Datacenter Core) (updated September 2019) 64 Bit Russian
2. Intel Cluster Studio for Linux Single Academic for 3 years_поставка 2011
3. Astra Linux Special Edition РУСБ.10015-01

12 Материально-техническое обеспечение дисциплины

1. Компьютер Intel i3/LCD22"/ AS H55M/2Gb/ HDD 500Gb/GF 1024Mb/DVD-RW/ИБП 800VA
2. Компьютер Intel Core i3 3220/DDR3 4Gb/Hdd 1Tb/GF 1Gb/DVD+RW/Sound Net km/LCD23"/
3. Компьютер" Intel Core i3/DDR 4GB/HDD 1 Tb/GF 1Gb/LCD23/ИБП"
4. Проектор Epson EB-W04LCD.WXGA 1280*800.3000:1.2800 ANSI Lumens