

Федеральное государственное бюджетное учреждение науки
Институт машиноведения им. А.А. Благонравова
Российской академии наук
(ИМАШ РАН)

УТВЕРЖДАЮ

Заместитель директора по вопросам
подготовки кадров высшей
квалификации ИМАШ РАН

к.ф.-м.н. Жданов Э.Р.

« 01 » _____ 2026 года

М.П.



Рабочая программа дисциплины
«Методы анализа больших данных (Big Data Analytics)»

Направление подготовки:	02.04.03 Математическое обеспечение и администрирование информационных систем
Профиль подготовки:	Искусственный интеллект и интеллектуальная обработка данных
Квалификация:	Магистр
Форма обучения	Очная

Москва, 2026 г.

1. Аннотация к дисциплине

1.1.	Составлена в соответствии с ФГОС ВО - магистратура по направлению подготовки 02.04.03 Математическое обеспечение и администрирование информационных систем, утвержденным приказом Минобрнауки от 23.08.2017г. № 812.
1.2.	Формирует у студентов системные знания о современных подходах, инструментах и технологиях обработки, хранения и анализа больших данных. Рассматриваются архитектурные особенности распределённых систем, методы предобработки и визуализации данных, а также алгоритмы машинного обучения, адаптированные для работы с большими объёмами информации.
1.3.	Особое внимание уделяется практическому применению инструментов экосистемы Hadoop/Spark, управлению жизненным циклом аналитических проектов и внедрению инновационных решений на основе больших данных в организациях.

2. Место дисциплины в структуре ООП

2.1.	Цикл (раздел) ООП: Б1.В.ОД.2
2.2.	1 курс, 1 семестр

3. Цель освоения дисциплин

3.1.	Формирование у обучающихся необходимых компетенций для успешного освоения образовательной программы
------	---

4. Задачи освоения дисциплины

4.1.	- сформировать знания о современных архитектурах и инфраструктуре для хранения и обработки больших данных, методах распределённых вычислений, этических и правовых аспектах работы с данными; - сформировать навыки работы с ключевыми инструментами Big Data (Hadoop, Spark, NoSQL), реализации пайплайнов обработки данных, применения методов машинного обучения к большим массивам; - сформировать умение проектировать и управлять аналитическими проектами, выбирать оптимальные методы и технологии для решения практических задач, интерпретировать результаты анализа и принимать обоснованные управленческие решения.
------	---

5. Перечень планируемых результатов обучения, соотнесенных с планируемыми результатами освоения образовательной программы

ПК-1	Создание информационных технологий нового поколения
ПК-1.1	Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации
ПК-1.2	Управление разработкой продуктов, услуг и решений на основе больших данных
ПК-1.3	Разработка и внедрение новых методов и технологий исследования больших данных

Код индикатора компетенции	Планируемые результаты обучения по дисциплине
ПК-1.1	Знать: - современные архитектуры и компоненты инфраструктуры Big Data; - этапы жизненного цикла аналитических систем; - методы мониторинга и оптимизации производительности распределённых систем. Уметь: - проектировать архитектуру хранилищ и вычислительных кластеров для задач анализа данных; - выбирать и внедрять инструменты для сбора, хранения и обработки данных; - организовывать процессы сопровождения и развития инфраструктуры. Иметь практический опыт: - развёртывания и администрирования кластера Hadoop/Spark; - настройки пайплайнов обработки данных;

	- мониторинга и оптимизации работы распределённых приложений.
ПК-1.2	<i>Знать:</i> - методологию управления проектами в области Big Data; - современные подходы к формированию продуктовых команд (Data Science, Data Engineering); - методы оценки эффективности аналитических продуктов. <i>Уметь:</i> - формулировать требования к аналитическим решениям; - организовывать процесс разработки и тестирования моделей на больших данных; - интегрировать аналитические сервисы в бизнес-процессы организации. <i>Иметь практический опыт:</i> - проектирования и реализации прототипа продукта на основе анализа больших данных; - презентации и защиты аналитического решения перед стейкхолдерами.
ПК-1.3	<i>Знать:</i> - современные методы машинного обучения и анализа данных для Big Data; - тенденции развития технологий искусственного интеллекта в области обработки больших объёмов информации; - принципы экспериментирования и внедрения инноваций. <i>Уметь:</i> - адаптировать существующие алгоритмы для распределённой реализации; - проводить исследовательские эксперименты с новыми методами анализа данных; - оценивать применимость новых технологий для решения практических задач. <i>Иметь практический опыт:</i> - реализации исследовательского проекта по анализу больших данных с использованием нестандартных методов или инструментов.

6. Тематический план

Наименование темы/раздела	Виды занятий и трудоемкость (в часах)					Формы контроля Форма ПА
	Всего	Л	ПЗ	ЛЗ	СР	
Тема 1. Введение в анализ больших данных	34	4	6	2	22	Отчет по практикуму по решению задач/10 Отчет по лабораторному практикуму /15
Тема 2. Методы предобработки и анализа данных	36	4	6	4	22	Отчет по практикуму по решению задач/10 Отчет по лабораторному практикуму /15
Тема 3. Машинное обучение на больших данных	38	4	8	4	22	Отчет по практикуму по решению задач/10 Отчет по лабораторному практикуму /15
Тема 4. Управление проектами и внедрение решений	36	2	8	4	22	Отчет по практикуму по решению задач/10 Отчет по лабораторному практикуму /15
Всего часов	144	14	28	14	88	100 ТКУ
Контроль, час	36					Экзамен
Объем дисциплины (в ак.ч.)	180					
Объем дисциплины (в з.е.)	5					

7. Содержание дисциплины

Наименование темы/раздела	Краткое содержание	Объем часов
Тема 1. Введение в анализ	Понятие Big Data: характеристики, источники, задачи.	4

больших данных	Архитектура распределённых систем. Обзор экосистем Hadoop и Spark. Хранение данных: HDFS, NoSQL-базы. Установка и настройка кластера Hadoop/Spark, загрузка и первичная обработка данных.	
Тема 2. Методы предобработки и анализа данных	Очистка, трансформация и интеграция данных. Методы выборки и агрегации. Визуализация больших данных. Реализация ETL-процесса с помощью Spark, построение интерактивных дашбордов.	4
Тема 3. Машинное обучение на больших данных	Особенности реализации алгоритмов ML в распределённой среде (MLlib). Классификация, регрессия, кластеризация, рекомендательные системы. Оценка моделей на больших выборках. Реализация задачи классификации или кластеризации на Spark MLlib.	4
Тема 4. Управление проектами и внедрение решений	Жизненный цикл проектов Big Data. Организация командной работы (DevOps для Data Science). Внедрение моделей в промышленную эксплуатацию. Проектирование и презентация прототипа аналитического решения для бизнес-кейса.	2

8. Самостоятельная работа студентов

Наименование темы/раздела	Виды самостоятельной работы	Трудоемкость в часах
Тема 1. Введение в анализ больших данных	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	22
Тема 2. Методы предобработки и анализа данных	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	22
Тема 3. Машинное обучение на больших данных	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	22
Тема 4. Управление проектами и внедрение решений	Работа с литературой, ЭБС, источниками в сети Internet, подготовка заданий для сдачи текущего контроля успеваемости	22

9. Тематика практических и лабораторных занятий

Наименование темы/раздела	Вид занятия	Краткое содержание занятия	Объем часов
Тема 1. Введение в анализ больших данных	практикум по решению задач	Решение задач	6
Тема 1. Введение в анализ больших данных	лабораторный практикум	Выполнение заданий в лабораторных условиях	2
Тема 2. Методы предобработки и анализа данных	практикум по решению задач	Решение задач	6
Тема 2. Методы предобработки и анализа данных	лабораторный практикум	Выполнение заданий в лабораторных условиях	4
Тема 3. Машинное обучение на больших данных	практикум по решению задач	Решение задач	8
Тема 3. Машинное обучение на больших данных	лабораторный практикум	Выполнение заданий в лабораторных условиях	4
Тема 4. Управление проектами и	практикум по	Решение задач	8

внедрение решений	решению задач		
Тема 4. Управление проектами и внедрение решений	лабораторный практикум	Выполнение заданий в лабораторных условиях	4

10. Фонд оценочных средств

12.1. Контрольные вопросы и задания для проведения текущего контроля
В приложении
12.2. Фонд оценочных средств для проведения промежуточной аттестации
В приложении
12.3. Порядок начисления баллов и критерии оценивания
В приложении

11. Учебно-методическое и информационное обеспечение дисциплины

9.1. Основная литература			
Авторы	Заглавие	Издательство, год	Эл. адрес
Волкова, В. Н.	Теория информационных процессов и систем: учебник и практикум для вузов / В. Н. Волкова. — 2-е изд., перераб. и доп.	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/583135
Иванов, И. В.	Теория информационных процессов и систем: учебник для вузов / И. В. Иванов. — 3-е изд., перераб. и доп.	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/585719
Е. В. Стельмашонок, В. Л. Стельмашонок, Л. А. Еникеева, С. А. Соколовская	Моделирование процессов и систем: учебник и практикум для вузов	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/583796
9.2. Дополнительная литература			
Авторы	Заглавие	Издательство, год	Эл. адрес
Краснова, Т. И	Английский язык для специалистов в области интернет-технологий. English for Internet Technologies: учебник для вузов / Т. И. Краснова, В. Н. Вичугов. — 3-е изд., перераб. и доп.	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/584047
Маркин, А. В.	Программирование на SQL: учебник и практикум для вузов / А. В. Маркин. — 3-е изд., перераб. и доп.	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/589589
Стружкин, Н. П.	Базы данных: проектирование: учебник для вузов / Н. П. Стружкин,	Москва: Издательство Юрайт, 2026	https://urait.ru/bcode/583031

12. Материально-техническое обеспечение дисциплины

<i>Учебные помещения</i>	кабинеты оснащены мебелью и оборудованием, соответствующими санитарным нормам и требованиям техники безопасности.
<i>Информационно-технологическая инфраструктура</i>	- серверы и хостинг: наличие сервера или облачного хранилища для размещения учебных ресурсов и платформ электронного обучения; - интернет-канал: стабильный широкополосный доступ к интернету для бесперебойной передачи больших объемов данных.
<i>Организационное обеспечение</i>	- технические средства обучения: компьютеры, ноутбуки, планшеты и мобильные устройства; - программы для видеоконференцсвязи для проведения лекций, консультаций и практических занятий в режиме реального времени; - антивирусное ПО и защита данных для предотвращения утечек конфиденциальной информации и защиты пользователей от киберугроз.
<i>Условия взаимодействия участников образовательного процесса</i>	- дистанционная поддержка преподавателей: помощь техническими специалистами настройки оборудования и программного обеспечения. - поддержка слушателя: организация технической поддержки, включая инструкции по установке и настройке необходимых приложений, устранение возникающих проблем.