МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГЕОДЕЗИИ И КАРТОГРАФИИ
(МИИГАиК)

Факультет геоинформатики и информационной безопасности
Кафедра информатики и геоинформационных технологий

**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 5BFEE1D286E9C95A0C0D2C611E434BD5
Владелец: Степанченко Алексей Леонидович
Действителен: с 23.09.2026 по 17.12.2027

«УТВЕРЖДАЮ»
Проректор
А. Л. Степанченко
18 апреля 2025 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ) «Б1.В.06 ТЕХНОЛОГИИ АНАЛИЗА БОЛЬШИХ ДАННЫХ»

Уровень высшего образования: магистратура

Направление подготовки: 09.04.03 Прикладная информатика

Направленность (профиль) подготовки: Анализ больших пространственных данных

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Год набора (приема на обучение): 2025

Срок получения образования: 2 года

Объем: в зачетных единицах: 5 з.е.
в академических часах: 180 ак.ч.

Москва, 2025

Разработчики:

Старший преподаватель кафедры Информатики и геоинформационных технологий Стрельцов А. С.

Доцент кафедры Информатики и геоинформационных технологий, к.н. Учаев Д. В.

Фонд оценочных средств по дисциплине (модулю) составлен в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 916

Согласование и утверждение

№	Подразделение	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	факультет Геоинформатики и информационной безопасности	Председатель ученого совета факультета	Матерухин А. В.	Рассмотрено	14.10.2024, № 10-24

1. Планируемые результаты обучения, соотнесенные с планируемыми результатами освоения образовательной программы

ПК-1 Способен выполнять работы по сбору, обработке и анализу больших пространственных данных

ПК-1.3 Анализирует большие пространственные данные

Знать:

ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных

ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных

Уметь:

ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных

ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности

Владеть:

ПК-1.3/Нв2 Навыками использования типовых алгоритмов анализа больших данных для решения практических задач профессиональной деятельности

ПК-1.3/Нв3 Навыками использования специализированного программного обеспечения для анализа больших данных

2. Показатели и критерии оценивания компетенций

Планируемые результаты освоения программы	Уровень овладения	Критерии оценивания (перечень сформированных ИДК/ЗУВ по уровням освоения)
ПК-1 Способен выполнять работы по сбору, обработке и анализу больших пространственных данных	Повышенный	ИДК ПК-1.3 Анализирует большие пространственные данные Знать: ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных Уметь: ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности Владеть: ПК-1.3/Нв2 Навыками использования типовых алгоритмов анализа больших данных для решения практических задач профессиональной деятельности ПК-1.3/Нв3 Навыками использования специализированного программного обеспечения для анализа больших данных

	Базовый	ИДК ПК-1.3 Анализирует большие пространственные данные Знать: ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных Уметь: ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности
	Пороговый	ИДК ПК-1.3 Анализирует большие пространственные данные Знать: ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных

3. Шкала оценивания

Оценка	Критерии оценивания (перечень сформированных ИДК/ЗУВ)
Отлично	ПК-1.3 Анализирует большие пространственные данные ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности ПК-1.3/Нв2 Навыками использования типовых алгоритмов анализа больших данных для решения практических задач профессиональной деятельности ПК-1.3/Нв3 Навыками использования специализированного программного обеспечения для анализа больших данных

Хорошо	ПК-1.3 Анализирует большие пространственные данные ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных ПК-1.3/Ум2 Использовать научную и научно-техническую литературу в области обработки больших данных при программной реализации алгоритмов анализа больших данных ПК-1.3/Ум3 Использовать применяемые в обработке больших данных программное обеспечение для решения аналитических и исследовательских задач профессиональной деятельности
Удовлетворительно	ПК-1.3 Анализирует большие пространственные данные ПК-1.3/Зн2 Технологии и алгоритмы анализа больших данных ПК-1.3/Зн3 Программные среды и языки, применяемые для обработки больших объемов данных
Неудовле-творительно	Результаты обучения не сформированы на достаточном уровне

4. Контрольные мероприятия по дисциплине

Вид контроля	Форма контроля/Оценочное средство
Текущий контроль	Практическая работа
Промежуточная аттестация	Экзамен

№ п/п	Наименование раздела	Контролируемые ИДК	Вид контроля/ используемые оценочные материалы	
			Текущий	Промежут. аттестация
1	Большие данные (Big Data). Введение в анализ больших данных	ПК-1.3	Практическая работа	Экзамен
2	Программное обеспечение в области анализа больших данных	ПК-1.3	Практическая работа	Экзамен
3	Технологии анализа больших объемов данных	ПК-1.3	Практическая работа	Экзамен

5. Порядок проведения промежуточной аттестации

Порядок проведения промежуточной аттестации регламентируется локальными нормативными актами Университета:

- Положением о формах, периодичности, порядке текущего контроля успеваемости и промежуточной аттестации обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.6 «Организация и проведение промежуточной аттестации».
- Положением о порядке осуществления образовательного процесса, проведения текущего контроля успеваемости и промежуточной аттестации обучающихся с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.8 «Проведение промежуточной аттестации с применением ЭО и ДОТ».

Промежуточная аттестация: Экзамен, 4 семестр.

Устанавливается следующая шкала оценивания уровней освоения компетенций.

Оценка «5» ставится, если студент:

1) полно и аргументированно отвечает по содержанию задания;

2) обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только по учебнику, но и самостоятельно составленные;

3) излагает материал последовательно и правильно.

Оценка «4» ставится, если студент дает ответ, удовлетворяющий тем же требованиям, что и для оценки «5», но допускает 1-2 ошибки, которые сам же исправляет.

Оценка «3» ставится, если студент обнаруживает знание и понимание основных положений данного задания, но:

1) излагает материал неполно и допускает неточности в определении понятий или формулировке правил;

2) не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры;

3) излагает материал непоследовательно и допускает ошибки.

Оценка «2» ставится, если студент обнаруживает незнание ответа на соответствующее задание, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал. Оценка «2» отмечает такие недостатки в подготовке студента, которые являются серьезным препятствием к успешному овладению последующим материалом.

6. Оценочные материалы текущего контроля

Раздел 1. Большие данные (Big Data). Введение в анализ больших данных

Контролируемые ИДК: ПК-1.3

Тема 1.2. Основы анализа, извлечения и представления данных

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Поиск и выявление больших данных (Big Data)

Провести поиск источников информации в сети Интернет: рассмотреть пор-тал открытых данных РФ. Определить можно ли классифицировать найденные данные как «Big Data». Сохранить данные. Подготовить отчет с выводами.

Раздел 2. Программное обеспечение в области анализа больших данных

Контролируемые ИДК: ПК-1.3

Тема 2.1. Платформы и инструменты анализа больших данных

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Описание структуры больших данных

Организовать импорт больших данных из RStudio в пакет Statistica. Описать структуру полученных данных. Провести описательную статистику в пакете Statistica. Подготовить отчет с выводами.

Раздел 3. Технологии анализа больших объемов данных

Контролируемые ИДК: ПК-1.3

Тема 3.1. Технологии KDD и Data Mining для анализа больших данных

Форма контроля/оценочное средство: Практическая работа

Тема 3.2. Ассоциативные правила и их применение для анализа больших данных.

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Поиск ассоциативных правил

По показателям больших данных, загруженных в RStudio, осуществить поиск ассоциативных правил. Подготовить отчет с выводами.

Тема 3.3. Алгоритмы кластеризации и их применение в анализе больших данных

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Кластеризация больших данных

Провести кластерный анализ больших данных, загруженных в RStudio. Применить метод k-средних и иерархическую классификацию. Выбрать оптимальный способ. Подготовить отчет с выводами.

Тема 3.4. Классификация больших данных

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Классификация больших данных с помощью деревьев решений

К показателям больших данных, загруженных в RStudio, применить алгоритм «Дерево решений». Описать структуру дерева. Выявить основные факторы, влияющие на классификацию. Подготовить отчет с выводами.

2. Классификация больших данных с помощью нейросети

Методами нейросетевой классификации определить возможность разбиения больших данных на кластеры, выделенные по показателям, характеризующим другую сферу явления. Подготовить отчет с выводами.

Тема 3.5. Корреляционный и регрессионный анализ больших данных. Линейная и логистическая регрессии

Форма контроля/оценочное средство: Практическая работа

Вопросы/Задания:

1. Прогнозирование с помощью линейной регрессии

Построить уравнение линейной регрессии по большим данным, загруженным в RStudio, средствами языка программирования R. Подготовить отчет с выводами.

7. Оценочные материалы промежуточной аттестации

4 семестр, Экзамен

Контролируемые ИДК: ПК-1.3

Вопросы/Задания:

1. Дополните предложение термином:

Одна из ключевых характеристик больших данных, наряду с объёмом (Volume) и скоростью поступления (Velocity), – _____ (Variety), отражающая разнородность форматов и источников данных.

2. Как называется этап предварительной обработки данных, устраняющий пропуски, дубликаты и явные ошибки в исходном наборе данных перед анализом?

3. Как называется статистическая мера, показывающая силу и направление линейной взаимосвязи между двумя количественными признаками, принимающая значения от −1 до 1?

4. Дополните предложение термином:

Метод _____ кластеризации последовательно объединяет (или разделяет) объекты в кластеры на основе их близости, формируя дерево вложенных кластеров (дендрограмму), в отличие от метода k-средних, заранее требующего указания числа кластеров.

5. Какие подходы применяются для описания структуры больших данных?

6. Дайте определение технологии KDD (Knowledge Discovery in Databases) и Data Mining, поясните их соотношение друг с другом и опишите роль Data Mining в общем жизненном цикле анализа данных.

7. Раскройте содержание понятия ассоциативных правил в интеллектуальном анализе данных: дайте определение, поясните смысл меры «достоверность» (confidence) правила и приведите пример практического применения такого анализа.

8. Опишите основные проблемы кластеризации больших данных: выбор числа кластеров, чувствительность к шуму и масштабируемость алгоритма – дайте определение каждой проблеме.

9. Опишите принципы классификации данных, дайте общее определение задачи классификации и объясните, почему один и тот же нейросетевой блок извлечения признаков (например, свёрточные слои) может применяться для классификации разных типов больших данных, отличающихся лишь конкретным набором признаков и целевых классов.

10. Опишите корреляционный и регрессионный анализ больших данных: дайте определение каждому виду анализа и укажите принципиальное различие в том, что именно каждый из них устанавливает между признаками.

11. Расположите этапы процесса KDD (обнаружения знаний в базах данных) в правильной последовательности.

1. Применение алгоритмов Data Mining для выявления закономерностей.
2. Отбор релевантных данных из исходного источника (базы данных, хранилища).
3. Предварительная обработка и очистка отобранных данных.
4. Преобразование данных к виду, пригодному для применения алгоритмов анализа.
5. Интерпретация и оценка полученных результатов.

12. Расположите этапы кластеризации больших данных методом k-средних в правильной последовательности.

1. Отнесение каждого объекта к ближайшему по расстоянию центру кластера.
2. Пересчёт положения каждого центра кластера как среднего значения всех отнесённых к нему объектов.
3. Случайный выбор начальных положений k центров кластеров.
4. Повторение отнесения объектов и пересчёта центров до тех пор, пока положения центров не перестанут существенно изменяться.

13. Расположите этапы построения и оценки модели классификации больших данных с использованием дерева решений в правильной последовательности.

1. Разделение данных на обучающую и тестовую выборки.
2. Оценка точности классификации по результатам на тестовой выборке.
3. Загрузка и предварительная обработка исходного набора данных.
4. Построение дерева решений на обучающей выборке путём последовательного выбора наиболее информативных признаков для разбиения.

5. Обрезка (pruning) дерева для снижения риска переобучения.
6. Получение предсказаний модели на тестовой выборке.

14. Расположите этапы построения уравнения линейной регрессии по большим данным в правильной последовательности.

1. Оценка параметров (коэффициентов) регрессии методом наименьших квадратов на обучающих данных.
2. Определение зависимой (целевой) переменной и независимых переменных-предикторов.
3. Проверка адекватности построенной модели (например, по значению коэффициента детерминации R^2).

15. Расположите этапы применения логистической регрессии для бинарной классификации в правильной последовательности.

1. Преобразование этой линейной комбинации через логистическую (сигмоидную) функцию в значение вероятности от 0 до 1.
2. Вычисление линейной комбинации значений признаков объекта с обученными весами модели.
3. Отнесение объекта к одному из двух классов по результату сравнения с порогом.
4. Сравнение полученной вероятности с пороговым значением (обычно 0, 5).

16. Установите соответствие между характеристикой больших данных («V») и её содержанием.

Volume (объём) - Разнородность форматов и источников данных
Velocity (скорость) - Физически большой размер накопленных данных
Variety (многообразие) - Неоднородное качество и надёжность данных из разных источников
Veracity (достоверность) - Высокая скорость поступления новых данных, часто в режиме, близком к реальному времени

17. Установите соответствие между этапом предварительной обработки данных и его содержанием.

Очистка (cleaning) - Объединение данных с более детального уровня в сводные показатели более общего уровня
Нормализация - Устранение пропусков, дубликатов и явных ошибок в данных
Агрегация - Приведение значений признаков к единому масштабу (диапазону)

18. Установите соответствие между нейросетевой архитектурой (или деревом решений) и типом данных, для которого она в первую очередь предназначена.

MLP (многослойный персептрон) - Последовательные данные, где важен порядок элементов (временные ряды, текст)
CNN - Универсальные табличные признаки без выраженной пространственной или временной структуры
RNN - Данные с выраженной пространственной локальной структурой (изображения)
LSTM - Табличные признаки, где важна интерпретируемость итогового правила классификации
Дерево решений - Длинные последовательности, где важны долгосрочные зависимости между удалёнными по времени элементами

19. Установите соответствие между понятием, связанным с ассоциативными правилами, и его определением.

Ассоциативное правило - Доля транзакций, содержащих одновременно A и B, среди всех транзакций набора данных
Поддержка (support) - Набор объектов, встречающийся совместно в транзакциях не реже заданного порога
Достоверность (confidence) - Закономерность вида «если присутствует набор A, то с определённой вероятностью присутствует и набор B»
Частое множество - Доля транзакций, содержащих и A, и B, среди всех транзакций, содержащих A

20. Установите соответствие между проблемой кластеризации больших данных и её содержанием.

Выбор числа кластеров - Классические алгоритмы вычислительно затратны и плохо справляются с действительно большими объёмами данных
Чувствительность к шуму - Многие алгоритмы требуют заранее указать число кластеров, хотя оно заранее неизвестно
Масштабируемость - Выбросы и ошибочные наблюдения могут существенно исказить результат кластеризации