



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И.
Носова»



УТВЕРЖДАЮ
Директор ИЭиАС
В.Р. Храмшин

25.04.2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление подготовки (специальность)
38.03.05 Бизнес-информатика

Направленность (профиль/специализация) программы
Аналитическая поддержка в управлении бизнес-процессами

Уровень высшего образования - бакалавриат

Форма обучения
очная

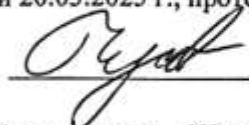
Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Бизнес-информатики и информационных технологий
Курс	3
Семестр	5

Магнитогорск
2025 год

Рабочая программа составлена на основе ФГОС ВО - бакалавриат по направлению подготовки 38.03.05 Бизнес-информатика (приказ Минобрнауки России от 29.07.2020 г. № 838)

Рабочая программа рассмотрена и одобрена на заседании кафедры Бизнес-информатики и информационных технологий 20.03.2025 г., протокол № 7

Зав. кафедрой



Г.Н. Чусавитина

Рабочая программа одобрена методической комиссией ИЭиАС 25.04.2025 г. протокол № 5

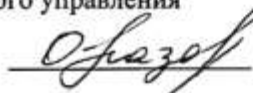
Председатель



В.Р. Храмшин

Согласовано:

Зав. кафедрой Менеджмента и государственного управления



О.Л. Назарова

Рабочая программа составлена:

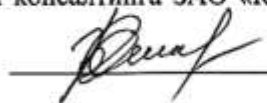
доцент кафедры БИиИТ, канд. пед. наук



Е.В. Карманова

Рецензент:

руководитель направления бизнес-анализа и консалтинга ЗАО «КонсОМ СКС»,
канд. техн. наук



В.А. Ошурков

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2026 - 2027 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2027 - 2028 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2028 - 2029 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2029 - 2030 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

1 Цели освоения дисциплины (модуля)

Целью освоения дисциплины является формирование у студентов практических навыков сбора, обработки и использования больших данных с применением методов машинного и глубокого обучения.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Анализ больших данных входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Информационные процессы, системы и сети

Основы алгоритмизации и программирования

Информатика

Экономическая статистика

Проектирование, разработка и оптимизация WEB-приложений

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Подготовка к сдаче и сдача государственного экзамена

Подготовка к защите и защита выпускной квалификационной работы

Цифровые платформы в бизнесе

Прикладной анализ данных в Python

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Анализ больших данных» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-1	Способен осуществлять архитектурные и реализационные решения по интеграции приложений информационных систем и облачных сервисов, а также планировать аналитические работы с использованием технологий больших данных
ПК-1.1	Осуществляет архитектурные и реализационные решения по интеграции приложений информационных систем и облачных сервисов
ПК-1.2	Решает профессиональные задачи по планированию аналитических работ с использованием технологий больших данных

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 5 зачетных единиц 180 акад. часов, в том числе:

- контактная работа – 94,1 акад. часов;
- аудиторная – 90 акад. часов;
- внеаудиторная – 4,1 акад. часов;
- самостоятельная работа – 50,2 акад. часов;
- в форме практической подготовки – 0 акад. час;
- подготовка к экзамену – 35,7 акад. час

Форма аттестации - экзамен

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Методология и технологии анализа больших данных								
1.1 Большие данные в прикладных задачах.	5	2		2	2	Проработка к контрольным вопросам	Устный опрос	ПК-1.2
1.2 Технологическая и нормативно-правовая основы обработки открытых и закрытых больших данных		2		2	4	Проработка теоретического материала, подготовка отчета по лабораторной работе	Устный опрос, проверка отчета по лабораторным работам	ПК-1.2
1.3 Направления исследований в области аналитики больших данных		2		2	4	Проработка теоретического материала, подготовка отчета по лабораторной работе	Устный опрос, проверка отчета по лабораторным работам	ПК-1.2
Итого по разделу		6		6	10			
2. Технологии сбора и предобработки больших данных								
2.1 Методы и инструменты автоматизированного сбора данных	5	2		4	4	Проработка теоретического материала, подготовка отчета по лабораторной работе	Устный опрос, проверка отчета по лабораторным работам	ПК-1.2
2.2 Основы предобработки больших		2		10	1	Проработка теоретического материала, подготовка отчета по лабораторной	Устный опрос, проверка отчета по лабораторным работам	ПК-1.2

						работе		
Итого по разделу		4		14	5			
3. Анализ слабоструктурированных данных социальных медиа								
3.1 Онтологическая модель пользователя социальной сети для задачи предиктивной аналитики	5	2		4	2,2	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.2
3.2 Методы сопоставления профилей пользователей в социальных сетях. Визуальный анализ данных		4		4	3	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.2
3.3 Методы анализа текстов в социальных сетях		3		8	8	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.2
Итого по разделу		9		16	13,2			
4. Технология распределенных вычислений больших данных								
4.1 Эволюция систем распределенного вычисления. Концепция Data Lake и pipelines	5	3		2	5	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.1
4.2 Архитектура Apache Hadoop. Файловая система HDFS.		6		6	1	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.2
4.3 Парадигма MapReduce		4		6	8	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.1
4.4 Операции обслуживания кластера Hadoop		4		4	8	Подготовка отчета по лабораторной работе	Проверка отчета по лабораторной работе	ПК-1.1
4.5 Экзамен								ПК-1.1, ПК-1.2
Итого по разделу		17		18	22			
Итого за семестр		36		54	50,2		экзамен	
Итого по дисциплине		36		54	50,2		экзамен	

5 Образовательные технологии

В ходе проведения занятий используются традиционные формы проведения занятий такие как:

1) информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

2) лабораторная работа – организация учебной работы с реальными материальными и информационными объектами.

На лекционных и лабораторных работах используются технологии проблемного обучения – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности студентов:

1) проблемная лекция – изложение материала, предполагающее постановку проблемных и дискуссионных вопросов, освещение различных научных подходов, авторские комментарии, связанные с различными моделями интерпретации изучаемого материала.

2) лабораторная работа на основе кейс-метода – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия производственной, общественной деятельности. Обучающиеся должны проанализировать ситуацию, разобраться в сути проблем, предложить возможные решения и выбрать лучшее из них. Кейсы базируются на реальном фактическом материале или же приближены к реальной ситуации.

Применяются интерактивные технологии – организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе личностно значимого для них образовательного результата. Интерактивность подразумевает субъект-субъектные отношения в ходе образовательного процесса и, как следствие, формирование саморазвивающейся информационно-ресурсной среды.

Формы учебных занятий, используемые в дисциплине, с использованием специализированных интерактивных технологий:

1. Лекция «обратной связи» – лекция–провокация (изложение материала с заранее запланированными ошибками), лекция-беседа, лекция-дискуссия, лекция пресс-конференция.

2. Лабораторная работа-дискуссия – коллективное обсуждение какого-либо спорного вопроса, проблемы, выявление мнений в группе (межгрупповой диалог, дискуссия как спор-диалог).

На лабораторных работах так же используются технологии проектного обучения, под которыми понимается организация образовательного процесса в соответствии с алгоритмом поэтапного решения проблемной задачи или выполнения учебного задания. Проект предполагает совместную учебно-познавательную деятельность группы студентов, направленную на выработку концепции, установление целей и задач, формулировку ожидаемых результатов, определение принципов и методик решения поставленных задач, планирование хода работы, поиск доступных и оптимальных ресурсов, поэтапную реализацию плана работы, презентацию результатов работы, их осмысление и рефлекссию.

При обучении используются информационно-коммуникационные образовательные технологии, под которыми понимается организация образовательного процесса, основанная на применении специализированных программных сред и технических средств работы с информацией.

На всех лекциях изложение содержания сопровождается компьютерными презентациями, содержащими текстовые, иллюстративные, графические и видеоматериалы.

На лабораторных работах и во время самостоятельной работы обучающиеся работают с ресурсами и сервисами образовательного портала <https://newlms.magtu.ru>

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины

а) Основная литература:

1. Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/145102> (дата обращения: 30.03.2025). — Режим доступа: для авториз. пользователей.-библиотечная система. — URL: <https://e.lanbook.com/book/162301>

2. Пальмов, С. В. Основы сбора и обработки больших данных : учебное пособие / С. В. Пальмов. — Самара : ПГУТИ, 2023. — 285 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/411830> (дата обращения: 30.03.2025). — Режим доступа: для авториз. пользователей.

б) Дополнительная литература:

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : УлГТУ, 2020. — 220 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/170653>

Кобзаренко, Д. Н. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес» : учебное пособие / Д. Н. Кобзаренко, А. Г. Мустафаев ; составитель Д. Н. Кобзаренко. — Махачкала : ДГУНХ, 2019. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/246542>

Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/145102>

Бродовская, Е. В. Большие данные в исследовании политических процессов : учебное пособие / Е. В. Бродовская, А. Ю. Домбровская. — Москва : МПГУ, 2018. — 88 с. — ISBN 978-5-4263-0712-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/125122>

Периодические издания:

1. Интеллектуальные системы [Электронный ресурс]. - Режим доступа: <http://intsysjournal.ru/>

2. Труды Института Системного Анализа РАН [Электронный ресурс]. - Режим доступа: <http://www.isa.ru/proceedings/>

3. Журнал «Программные продукты и системы» - <http://www.swsys.ru/>.

в) Методические указания:

1. Гаврилова И.В. Контрольно-измерительные материалы по дисциплине «Интеллектуальные информационные системы» для обучающихся направления подготовки 09.03.03 «Прикладная информатика» всех форм обучения. Магнитогорск: МГТУ им. Г.И. Носова, 2015. 37 с.

2. Курзаева Л.В., Гаврилова И.В. Методические указания для оценки знаний по дисциплинам «Системы поддержки принятия решений», «Экспертные системы и системы поддержки принятия решений» для обучающихся направления подготовки 09.03.03 «Прикладная информатика», 38.03.05 «Бизнес-информатика» всех форм обучения. Магнитогорск: МГТУ им. Г.И. Носова, 2015. 18 с.

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
7Zip	свободно распространяемое ПО	бессрочно
Anaconda Python	свободно распространяемое ПО	бессрочно
LibreOffice	свободно распространяемое ПО	бессрочно
Браузер Mozilla Firefox	свободно распространяемое ПО	бессрочно
Браузер Yandex	свободно распространяемое ПО	бессрочно
MS Visual Studio Code	свободно распространяемое ПО	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: https://elibrary.ru/project_risc.asp

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Материально-техническое обеспечение дисциплины включает:

Учебные аудитории для проведения занятий лекционного типа Персональный компьютер (или ноутбук) с пакетом Office, с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Доска, мультимедийный проектор, экран. Мультимедийные презентации к лекциям, учебно-наглядные пособия

Учебные аудитории для проведения лабораторных занятий Персональные компьютеры с пакетом Office; с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер MozillaFirefox.

Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки Персональные компьютеры с пакетом Office; с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер MozillaFirefox.

Аудитории для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации Персональные компьютеры с пакетом Office и выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер MozillaFirefox.

Аудитория для хранения и профилактического обслуживания учебного оборудования Мебель для хранения и обслуживания оборудования (шкафы, столы), учебно-методические материалы, компьютеры, ноутбуки, принтеры

Учебно-методическое обеспечение самостоятельной работы обучающихся
Перечень тем для подготовки к занятиям:

№ занятия	Краткое содержание занятия, семинара
1	Большие данные. Структуры и виды данных, больших данных в экономике. Задание: провести обзор он-лайн сервисов предоставляющих открытые датасеты (представить описание структур найденных датасетов). <i>Полезные источники:</i> – Ресурс с большими данными (хранилище открытых датасетов правительства США.) - https://data.gov/ – Ресурс с большими данными - https://archive-beta.ics.uci.edu/ – Ресурс с большими данными (Датасеты NASA, содержащие информацию об атмосфере Земли, океанах, криосфере, солнечных вспышках) - https://earthdata.nasa.gov/ – Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165835 (дата обращения: 15.06.2022). — Режим доступа: для авториз. пользователей.
2	Технологическая и нормативно-правовая основы обработки открытых и закрытых больших данных. Задание: изучить законодательство в области обработки данных. Какие регламентирующие документы относятся к области обработки больших данных в экономике? <i>Полезные источники:</i> Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165835 (дата обращения: 15.06.2022). — Режим доступа: для авториз. пользователей.
3	Направления исследований в области аналитики больших данных. Методы, инструменты анализа больших данных. Управление экономикой на основе данных. Задание: подготовить обзор современных направлений в области аналитики больших данных; сформулировать перспективную тему исследований в данной области;

	выделить задачи, которые необходимо решить в рамках выбранной темы. <i>Полезные источники:</i> – Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165835 (дата обращения: 15.06.2022). — Режим доступа: для авториз. пользователей. – Data Engineering: ETL, ELT, Data Pipeline, Data Warehouse, Data Lakes, Data Marts - https://ivan-shamaev.ru/data-engineering-etl-pipeline-data-warehouse-datalake/
4	Облачные вычисления и большие данные Задание: выполните сравнительный анализ инструментов обработки больших и малых данных. <i>Полезные источники:</i> Методы и модели исследования сложных систем и обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагина и А. Д. Хомоненко. — Санкт-Петербург : Лань, 2020. — 236 с. — ISBN 978-5-8114-4006-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/126938 (дата обращения: 06.07.2022). — Режим доступа: для авториз. пользователей.
5	Методы и инструменты автоматизированного сбора больших данных. Web-scraping, особенности работы с API. Приемы агрегирования данных. Построение дашбордов на основе больших данных. Задание: реализуйте web-scraping социальной сети - получите данные с постами в группах социальной сети Вконтакте. Сформируйте датасет, содержащий следующую информацию (дата поста, тема поста, содержание поста, количество репостов, количество “лайков”, комментарии к посту). <i>Полезные источники:</i> Талипов, Н. Г. Технологии интеллектуального анализа данных : учебное пособие / Н. Г. Талипов. — Казань : КНИТУ-КАИ, 2020. — 308 с. — ISBN 978-5-7579-2488-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/193530 (дата обращения: 11.06.2022). — Режим доступа: для авториз. пользователей.
6	Основы предобработки больших данных. Очистка данных. Устранение "выбросов". Анализ атрибутов и описательных статистик для больших данных. Анализ и обнаружение взаимосвязей в больших данных. Современные метрики проектов (экономическая часть). Задание: дан датасет социально-экономические показатели Чикаго. Необходимо проанализировать датасет на предмет выбросов, какие поля содержат нулевые

	значения, некорректные типы данных. Оцените имеется ли корреляция между отдельными показателями датасета. Какие прогнозируемые метрики можно предложить к данному датасету? <i>Полезные источники:</i> – Три способа обнаружения выбросов - История данных Колина Горри (англ.) - http://colingorrie.github.io/outlier-detection.html – Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2021. — 188 с. — ISBN 978-5-8114-6810-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165835 (дата обращения: 15.06.2022). — Режим доступа: для авториз. пользователей.
7	Факторный анализ. Задание: дан датасет социально-экономические показатели Чикаго. Необходимо провести факторный анализ, выделив, какие факторы влияют на социально-экономические показатели города. <i>Полезные источники:</i> Талипов, Н. Г. Технологии интеллектуального анализа данных : учебное пособие / Н. Г. Талипов. — Казань : КНИТУ-КАИ, 2020. — 308 с. — ISBN 978-5-7579-2488-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/193530 (дата обращения: 11.06.2022). — Режим доступа: для авториз. пользователей.
8	Онтологическая модель пользователя социальной сети для задачи предиктивной аналитики Статистическое представление пользователя в социальной сети, социальный портрет пользователя, метаинформация, социальный граф пользователя, социальная сеть ВКонтакте. Задание: 1. Используя методы VK API получить данные о 3 пользователях VK: – Количество сообществ, в которое входит пользователь – Количество друзей пользователя соц. сети – Количество подписчиков у пользователя – Количество подписок у пользователя – Количество постов пользователя – Количество комментариев пользователя 2. Построить один социальный граф пользователя, содержащий идентификатор анализируемого пользователя сети с множеством его непосредственных друзей. 3. Представить для одного пользователя его социальный портрет: общая информация о пользователе; статистические данные; интересы пользователей и друзей пользователя; социальный граф. Визуализировать полученные данные с помощью он-лайн сервисов интерактивной инфографики. <i>Полезные источники:</i> Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных

	больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : УлГТУ, 2020. — 220 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/170653 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
9	Методы сопоставления профилей пользователей в социальных сетях Задание: 1. Изучить возможность сервиса https://search4faces.com/ Провести анализ точности поиска людей по фото (с дополнительной информацией и нет), оформить чек лист для 20 фотографий, дать оценку полученным результатам. 2. Получить данные с профилями пользователей VK (не менее 100 пользователей), построить датасет профилей (ссылка на профиль, ФИО, дата рождения, место проживания, место рождения, список друзей, список постов, данные о месте работы (предыдущих местах работы), данные о месте учебы (предыдущих местах), список контактов пользователя, ссылки на другие сайты, телефон, аватар профиля (последние 5 фотографии из профиля)). 3. Разработать приложение для поиска схожих профилей по загруженному пользователю. Рейтинг включает в себя следующие критерии: наличие схожих фотографий, наличие схожих контактов, наличие схожего места работы, наличие схожего места учебы, наличие схожих постов, наличие схожих друзей. <i>Полезные источники:</i> Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : УлГТУ, 2020. — 220 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/170653 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
10	Методы анализа текстов в социальных сетях Смысл текстов, эмоциональная окраска, тональность Задание: Получить датасет с комментариями пользователей для любой открытой группы VK (количество комментариев не менее 1 тыс.). Построить модели классификации эмоциональной окраски комментариев: позитивность, негативность, нейтральность с указанием коэффициента тональности (от 0 до 1). Сравнить полученные модели по точности предсказания. <i>Полезные источники:</i> Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : УлГТУ, 2020. — 220 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/170653 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
11	Эволюция систем распределенного вычисления. Концепция Data Lake и

	pipelines Задание. 1. Провести обзор свободно-распространяемых библиотек и фреймворков для реализации распределенных вычислений на больших данных <i>Полезные источники:</i> Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
12	Архитектура Apache Hadoop. Файловая система HDFS. Задание: 1. Установите Java SE Development Kit. Скачайте и разверните дистрибутив Hadoop с официального сайта Apache. Настройте переменные окружения. Изучите конфигурационные параметры в документации Hadoop. <i>Полезные источники:</i> Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей. Методы и модели исследования сложных систем и обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагина и А. Д. Хомоненко. — Санкт-Петербург : Лань, 2020. — 236 с. — ISBN 978-5-8114-4006-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/126938 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
13	Парадигма MapReduce Задание. Реализуйте задачу по анализу текстов в парадигме MapReduce. <i>Полезные источники:</i> Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
14	Операции обслуживания кластера Hadoop Задание. 1. Реализуйте и продемонстрируйте в виде отчета следующие сервисные

	операции для кластера Hadoop относительно узла данных: – вывод из эксплуатации; – повторное включение в кластер; – перевод в состояние обслуживания; – вывод из состояния обслуживания. <i>Полезные источники:</i> Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.
15	Оптимизация и управление ресурсами Задание: 1. Изучите пример реализации кастомного сервиса, который следит за корректной утилизацией yarn очереди на Hadoop кластере - https://artydev.ru/posts/queue-hadoop-cluster/. Укажите области оптимизации и управления ресурсами в Hadoop. <i>Полезные источники:</i> Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 07.07.2022). — Режим доступа: для авториз. пользователей.

7 Оценочные средства для проведения промежуточной аттестации по дисциплине «Анализ больших данных»

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК-1: Способен осуществлять архитектурные и реализационные решения по интеграции приложений информационных систем и облачных сервисов, а также планировать аналитические работы с использованием технологий больших данных		
ПК – 1.1	Осуществляет архитектурные и реализационные решения по интеграции приложений информационных систем и облачных сервисов	Теоретические вопросы (к экзамену, зачету): 1. Методы сбора требований к системе обработки больших данных 2. Нормативно-правовые аспекты к системам хранения и обработки данных 3. Критерии оценки качества требований к Big Data-системам 4. Архитектурные требования к распределенным системам (например, Hadoop-кластерам) 5. Компоненты технико-экономического обоснования (ТЭО) для проекта в области больших данных 6. Оценка экономической эффективности внедрения Hadoop-кластера по сравнению с облачными решениями (AWS, Google Cloud) 7. Критерии выбора между open-source (Apache Spark) и коммерческими (SAS, IBM) инструментами аналитики 8. Учет затрат на предобработку данных (ETL) в ТЭ. Приведите пример расчета для датасета в 1 ТБ. 9. Риски (технические, финансовые, юридические) для системы анализа соцсетей Практические задания: 1. Анализ требований к системе сбора данных из социальных сетей Цель: Научиться выявлять и документировать функциональные и нефункциональные требования к системе. Задача: 1. Выберите конкретную задачу (например, мониторинг общественного мнения по продукту в VK). 2. Проведите анализ стейкхолдеров (кто заинтересован в системе: маркетологи, аналитики, юристы). Сформулируйте: 3. 5 функциональных требований (например, "Система должна собирать посты с ключевыми словами"). 4. 3 нефункциональных требования (например, "Обработка данных должна занимать не более 1 часа"). 5. Оформите требования в виде User Stories или SRS-документа (шаблон прилагается). 2. Документирование требований к подсистеме предобработки данных Цель: Разработать спецификацию для ETL-процесса (Extract, Transform, Load). Задача: Даны "сырые" данные из соцсетей (CSV-файл с полями: текст, дата, автор, лайки). Выявите требования к этапам:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		1. Очистка (удаление дубликатов, обработка пропусков). 2. Трансформация (нормализация текста, выделение тональности). 3. Загрузка (в каком формате данные должны сохраняться?). 4. Создайте схему процесса в виде блок-схемы (например, в Draw.io). 5. Укажите, какие инструменты (Python/Pandas, Apache NiFi) будут использоваться и почему. 3. Расчет ROI для внедрения системы предиктивной аналитики Цель: Научиться обосновывать экономическую целесообразность проекта. Задача: Компания хочет внедрить систему прогнозирования спроса на основе данных соцсетей. Дано: Затраты: Разработка: 500 тыс. руб. Nadoor-кластер: 200 тыс. руб./год. Обслуживание: 100 тыс. руб./год. Ожидаемая выгода: Снижение логистических издержек: 300 тыс. руб./год. Рост продаж: 400 тыс. руб./год. • Рассчитайте ROI за 3 года. • Определите срок окупаемости (Payback Period). • Проведите анализ чувствительности: как изменится ROI, если выгода окажется на 20% ниже? 4. Разработка ТЗ для подсистемы сбора данных Цель: Сформировать техническое задание на подсистему сбора данных из Twitter. Задача: На основе требований из предыдущих заданий составьте ТЗ, включив: • Цель системы (например, мониторинг упоминаний бренда). • Функциональные требования (частота сбора, фильтрация спама). • Технические требования (API VK, хранение в HDFS). • Экономические показатели (бюджет на разработку — 150 тыс. руб.). • Оформите документ по ГОСТ 34.602-89 (шаблон прилагается). Комплексное задание: Тема: Разработка требований к системе анализа Big Data для интернет-магазина. Этапы:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		Анализ потребностей: 1. Интервью с заказчиком (роли: менеджер, аналитик, DevOps). 2. Выявление ключевых задач (прогнозирование спроса, кластеризация клиентов). Документирование: 3. Функциональные требования (например, интеграция с 1С, генерация отчетов). 4. Нефункциональные (масштабируемость, время отклика < 2 сек). 5. Ограничения (бюджет, сроки, совместимость с Hadoop). Прототипирование: 6. Создайте макет интерфейса аналитика (Figma/Mockup) с визуализацией данных. 7. Напишите SQL-запрос для агрегации данных (например, средний чек по регионам). Итоговый результат: SRS-документ (шаблон ГОСТ 34.602-89). Презентация для стейкхолдеров (5-7 слайдов).
ПК – 1.2	Решает профессиональные задачи по планированию аналитические работы с использованием технологий больших данных	Теоретические вопросы (к экзамену, зачету): 1. Характеристики больших данных. 2. Технология факторного анализа больших данных. 3. Возможности методов интеллектуального анализа данных в экономике. Большие данные в экономике. 4. Законодательство в области больших данных. 5. Платформы бизнес-аналитики для анализа образовательных данных. 6. Этапы аналитического проекта при анализе образовательных данных. 7. Инструменты построения концептуальных карт 8. Методы сбора данных. Web-scraping, особенности работы с API. 9. Методы обнаружения взаимосвязей в больших данных. 10. Технология предобработки данных. 11. Методы предобработки данных. Очистка данных. 12. Методы предобработки данных. Сокращение данных. 13. Методы предобработки данных. Трансформация данных. 14. Методы предобработки данных. Интеграция данных. Практические задания: 1. Построить с использованием сервиса Yandex DataLens инфопанель (дашборд) для визуализации параметров зависимости стоимости квартир от отдельных показателей (для разных районов города). 2. Реализуйте web-scraping социальной сети - получите данные с постами в группах социальной сети Вконтакте. Сформируйте датасет, содержащий следующую информацию (дата поста, тема поста,

<i>Код индикатора</i>	<i>Индикатор достижения компетенции</i>	<i>Оценочные средства</i>
		содержание поста, количество репостов, количество “лайков”, комментарии к посту). Постройте визуализацию полученного датасета. 3. Получить данные с профилями пользователей VK (не менее 100 пользователей), построить датасет профилей (ссылка на профиль, ФИО, дата рождения, место проживания, место рождения, список друзей, список постов, данные о месте работы (предыдущих местах работы), данные о месте учебы (предыдущих местах), список контактов пользователя, ссылки на другие сайты, телефон, аватар профиля (последние 5 фотографии из профиля). 4. Реализовать метод PCA факторного анализа для определения важных признаков датасета, участвующих в предсказании стоимости квартир на рынке недвижимости. Комплексное задание: Вам дан датасет по стоимости квартир в городе Магнитогорске. Необходимо: 1. Провести факторный анализ, выделив, какие факторы влияют на формирование стоимости квартир (однокомнатных, двухкомнатных, трехкомнатных). 2. Разбить множество на классы, используя метод бинарной классификации. Критерии разбиения подобрать самостоятельно. 3. Построить ROC-кривую и написать пояснение к ней.

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Анализ больших данных» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме экзамена.

Экзамен по данной дисциплине проводится в устной форме по экзаменационным билетам, каждый из которых включает 2 теоретических вопроса и одно практическое задание.

Показатели и критерии оценивания экзамена:

– на оценку **«отлично»** (5 баллов) – обучающийся демонстрирует высокий уровень сформированности компетенций, всестороннее, систематическое и глубокое знание учебного материала, свободно выполняет практические задания, свободно оперирует знаниями, умениями, применяет их в ситуациях повышенной сложности.

– на оценку **«хорошо»** (4 балла) – обучающийся демонстрирует средний уровень сформированности компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.

– на оценку **«удовлетворительно»** (3 балла) – обучающийся демонстрирует пороговый уровень сформированности компетенций: в ходе

контрольных мероприятий допускаются ошибки, проявляется отсутствие отдельных знаний, умений, навыков, обучающийся испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

- на оценку **«неудовлетворительно»** (2 балла) – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.

- на оценку **«неудовлетворительно»** (1 балл) – обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач.