

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гнатюк Сергей Иванович
Должность: Первый проректор
Дата подписания: 06.10.2025 10:09:11
Уникальный программный ключ:
5ede28fe5b714e680817c5c132d4ba793a6b4421

Министерство сельского хозяйства Российской Федерации

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ К.Е. ВОРОШИЛОВА»**

«Утверждаю»
Декан факультета экономики и
управления АПК

Шевченко М.Н. _____
«25» апреля 2025 г.

РАБОЧАЯ ПРОГРАММА

учебной дисциплины «Основы работы с Большими данными»
для направления подготовки 38.03.05 Бизнес-информатика
направленность (профиль) Бизнес-информатика

Год начала подготовки – 2025

Квалификация выпускника – бакалавр

Луганск, 2025

Рабочая программа составлена с учетом требований:

- порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 06.04.2021 № 245 (с изменениями и дополнениями);
- федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 38.03.05 Бизнес- информатика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 29.07.2020г. № 838 (с изменениями и дополнениями).

Преподаватели, подготовившие рабочую программу:

старший преподаватель _____ **Ю.А. Горячкова**
кафедры информационных технологий,
математики и физики

Рабочая программа рассмотрена на заседании кафедры информационных технологий, математики и физики (протокол № 8 от «07» апреля 2025 г.).

Заведующий кафедрой _____ **В.Ю. Ильин**

Рабочая программа рекомендована к использованию в учебном процессе методической комиссией факультета экономики и управления АПК (протокол № 8 от «24» апреля 2025 г.).

Председатель методической комиссии _____ **А.В. Худолей**

Руководитель основной профессиональной образовательной программы _____ **В.Ю. Ильин**

1. Предмет. Цели и задачи дисциплины, её место в структуре образовательной программы

Целью дисциплины является ознакомление студентов с ключевыми понятиями и основными направлениями анализа больших данных, современными мировыми трендами развития информационных технологий в данной области; формирование базовых компетенций по сбору, обработке и визуализации массовых данных для выполнения профессиональных задач.

Исходя из цели, в процессе изучения дисциплины решаются следующие **задачи**:

- усвоение основных понятий, категорий, терминов и определений, относящихся к области работы с большими данными;
- изучение специфики процессов информационной деятельности в области больших данных;
- формирование культуры работы с информацией;
- приобретение навыков визуализации данных;
- ознакомление с методами анализа информационных массивов, с технологиями обработки и хранения больших данных;
- выработка навыков использования информационных ресурсов для решения прикладных задач;
- повышение качества выполняемых выпускных работ и формирование компетенций, направленных на проведение самостоятельных научных исследований.

Место дисциплины в структуре образовательной программы.

Дисциплина «Основы работы с Большими данными» относится к части, формируемой участниками образовательных отношений (Б1.В.08) основой профессиональной образовательной программы высшего образования (далее – ОПОП ВО).

Данная дисциплина базируется на знаниях, полученных при изучении предметов: «Современные информационные технологии и системы искусственного интеллекта», «Базы данных», «Математический анализ».

Требования к предварительной подготовке обучающихся:

знать:

- базовые понятия информационных технологий;
- принципы обработки информации;
- общие принципы работы информационно-коммуникационных технологий;

уметь:

- работать с прикладными программами общего назначения;

использовать:

- телекоммуникационные технологии для решения учебных и профессиональных задач.

Освоение данной дисциплины необходимо для выполнения задания производственной практики и выпускной квалификационной работы.

Последующие читаемые дисциплины: «Объектно-ориентированный анализ», «OLAP&OLTP системы обработки и анализа данных».

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

Коды компетенций	Формулировка компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
ПК-3	Способен разрабатывать модели бизнес-процессов заказчика с учетом требований к информационным системам	ПК-3.1. Составляет анкеты и программы интервью с предполагаемыми представителями заказчика и будущими пользователями системы, анализирует исходную документацию, описывает автоматизируемые с помощью информационных систем бизнес-процессы, формирует требования к компонентам информационной системы	Знать: основы информационной и библиографической культуры работы с большими данными уметь: применять методы, приемы и способы решения стандартных задач профессиональной деятельности при работе с массивами больших данных иметь навыки: использования на практике информационно-коммуникационных технологий, с учетом основных требований информационной безопасности, для сбора, обработки и анализа больших данных
		ПК-3.2. Знает и применяет инструменты, методики описания и моделирования бизнес-процессов, осуществляет разработку моделей	Знать: основные понятия термина Большие Данные; основные понятия Data Mining; основные способы хранения и анализа данных; методы извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети). уметь: анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия иметь навыки: хранения и анализа данных; навыками извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети)

3. Объём дисциплины и виды учебной работы

Виды работ	Очная форма обучения		Заочная форма обучения	Очно-заочная форма обучения
	всего	в т.ч. по семестрам	всего часов	всего часов
		5 семестр	–	7 семестр
Общая трудоёмкость дисциплины, зач.ед./часов, в том числе:	6/216	6/216	–	6/216
Контактная работа, часов:	72	72	–	44
- лекции	36	36	–	22
- практические (семинарские) занятия	36	36	–	22
- лабораторные работы	–	–	–	–
Самостоятельная работа, часов	117	117	–	172
Контроль, часов	27	27	–	–
Вид промежуточной аттестации (зачёт, экзамен)	экзамен	экзамен	–	экзамен

4. Содержание дисциплины

4.1. Разделы дисциплины и виды занятий (тематический план)

Раздел дисциплины (тема)		Л	ПЗ	ЛР	СРС
Очная форма обучения					
Тема 1. Введение в большие данные		2	2	–	18
Тема 2. Жизненный цикл аналитики данных		4	4	–	18
Тема 3. Высокопроизводительные вычисления		6	6	–	18
Тема 4. Масштабирование и многоуровневое хранение данных		6	6	–	18
Тема 5. Визуализация данных и результатов анализа		4	4	–	18
Тема 6. Классификация задач анализа данных		4	4	–	18
Тема 7. Сложные методы аналитики		6	6	–	18
Тема 8. Основы построения нейронных сетей для анализа данных		4	4		18
Всего		36	36	–	144
Заочная форма обучения					
–		–	–	–	–
Очно-заочная форма обучения					
1	Тема 1. Введение в большие данные	2	2	–	20
2	Тема 2. Жизненный цикл аналитики данных	2	2	–	20
3	Тема 3. Высокопроизводительные вычисления	4	4	–	22
4	Тема 4. Масштабирование и многоуровневое хранение данных	4	4	–	22
5	Тема 5. Визуализация данных и результатов анализа	2	2	–	22

6	Тема 6. Классификация задач анализа данных	2	2	–	22
7	Тема 7. Сложные методы аналитики	4	4	–	22
8	Тема 8. Основы построения нейронных сетей для анализа данных	2	2	–	22
Всего		22	22	–	172

4.2. Содержание разделов учебной дисциплины

Тема 1. Введение в большие данные.

История и причины появления термина Big Data (Большие данные). Характеристики и источники Big Data. Четыре основных типа данных. Аналитика данных. Задачи, решаемые Big Data.

Тема 2. Жизненный цикл аналитики данных.

Методы и инструменты BusinessIntelligence (BI). ETL(Extract, Transform, Load)-процесс. Средства BI. Инструменты анализа BI: средства предоставления информации, средства интеграции, средства анализа. Online Analytical Processing (OLAP) – оперативная аналитическая обработка данных. Понятие жизненного цикла аналитики данных.

Тема 3. Высокопроизводительные вычисления.

История проекта Hadoop и MapReduce. Hadoop Distributed File System. Архитектура Hadoop. Технология Map Reduce.

Тема 4. Масштабирование и многоуровневое хранение данных

NoSQL(NotOnlySQL). Масштабируемость. Репликация. CAP теорема. MongoDB.

Тема 5. Визуализация данных и результатов анализа

Типы, задачи и виды визуализации. Графики, диаграммы, инфографика, Интерактивный сторителлинг, дашборды. Язык R и его возможности. AmazonS3. Дедупликация данных.

Тема 6. Классификация задач анализа данных

DataMining. Интеллектуальный анализ данных, его, отличия и задачи. Text Mining. Web Mining. Web Content Mining. Web Usage Mining. Social media mining. RapidMiner.

Тема 7. Сложные методы аналитики

Статистические гипотезы и критерии. Машинное обучение. Метрический и линейных классификаторы. ROC–кривая. Кластерный анализ.

Тема 8. Основы построения нейронных сетей для анализа данных

Основы нейронных сетей. Обучение нейронной сети.

Создание нейросети в библиотеке Keras: подготовка данных; конфигурация нейронной сети; настройки. Функция потерь. Алгоритм оптимизации.

4.3. Перечень тем лекций

№ п/п	Тема лекции	Объём, ч		
		форма обучения		
		очная	заочная	очно- заочная
1.	Тема лекционного занятия 1. Введение в большие данные	2	–	2
2.	Тема лекционного занятия 2. Жизненный цикл аналитики данных	4	–	2
3.	Тема лекционного занятия 3. Высокопроизводительные вычисления	6	–	4
4.	Тема лекционного занятия 4. Масштабирование и многоуровневое хранение данных	6	–	4
5.	Тема лекционного занятия 5. Визуализация данных и	4	–	2

	результатов анализа			
6.	Тема лекционного занятия 6. Классификация задач анализа данных	4	–	2
7.	Тема лекционного занятия 7. Сложные методы аналитики	6	–	4
8.	Тема лекционного занятия 8. Основы построения нейронных сетей для анализа данных	4	–	2
Всего		36	–	22

4.4. Перечень тем практических (семинарских) занятий

№ п/п	Тема практического (семинарского) занятия	Объём, ч		
		форма обучения		
		очная	заочная	очно-заочная
1	Тема практического занятия 1. Основы работы с аналитической платформой DEDUCTOR	4	–	2
2	Тема практического занятия 2. Применение факторного и корреляционного анализа	4	–	2
3	Тема практического занятия 3. Разбиение данных, квантования и фильтрации для трансформации данных	4	–	4
4	Тема практического занятия 4. Использование стандартных математических функций при анализе и формировке данных	4	–	2
5	Тема практического занятия 5. Ассоциативные правила и использовать визуализаторы «Популярные наборы», «Правила», «Дерево правил», «Что-если».	6	–	4
6	Тема практического занятия 6. Применение методов Data Mining для решения задач прогнозирования временных рядов на примере построения модели	6	–	4
7.	Тема практического занятия 7. Применение обработчика «скрипт» для решения задач прогнозирования на примере прогноза продаж	4	–	2
8.	Тема практического занятия 8. решения задач кластеризации с применением карт Кохонена	4	–	2
Всего		36	–	22

4.5. Перечень тем лабораторных работ.

Не предусмотрены.

4.6. Виды самостоятельной работы студентов и перечень учебно-методического обеспечения для самостоятельной работы обучающихся

4.6.1. Подготовка к аудиторным занятиям

Материалы лекций являются основой для изучения теоретической части дисциплины и подготовки студента к практическим занятиям.

При подготовке к аудиторным занятиям студент должен:

- изучить рекомендуемую литературу;
- просмотреть самостоятельно дополнительную литературу по изучаемой теме.

Основной целью практических занятий является изучение отдельных наиболее сложных и интересных вопросов в рамках темы, а также контроль за степенью усвоения пройденного материала и ходом выполнения студентами самостоятельной работы.

4.6.2. Перечень тем курсовых работ (проектов)

Не предусмотрены.

4.6.3. Перечень тем рефератов, расчетно-графических работ и иных видов индивидуальных работ

Не предусмотрены.

4.6.4. Перечень тем и учебно-методического обеспечения для самостоятельной работы обучающихся

№ п/п	Тема самостоятельной работы	Учебно-методическое обеспечение	Объем, ч		
			форма обучения		
			очная	заочная	очно-заочная
1.	Введение в большие данные	1. Параскевов, А. В. Большие данные : учебное пособие / А. В. Параскевов, А. Э. Сергеев. - Москва ; Вологда : Инфра-Инженерия, 2024. - 148 с. - ISBN 978-5-9729-2120-1. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2169699 (дата обращения: 06.03.2025). – Режим доступа: по подписке.	18	–	20
2.	Жизненный цикл аналитики данных	2. Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. - Москва : Изд. Дом МИСиС, 2016. - 52 с. - ISBN 978-5-87623-969-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/1232190 (дата обращения: 06.03.2025). – Режим доступа: по подписке.	18	–	20
3.	Высокопроизводительные вычисления	3. Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов ; Министерство науки и высшего образования Российской Федерации, Национальный исследовательский Московский государственный строительный университет, кафедра информационных систем, технологий и автоматизации в строительстве. - Москва : Издательство МИСИ - МГСУ,	18	–	22
4.	Масштабирование и управление многоуровневое хранение данных		18	–	22
5.	Визуализация данных и результатов анализа		18	–	22
6.	Классификация задач анализа данных		18	–	22
7.	Сложные методы аналитики		18	–	22
8.	Основы построения нейронных сетей для анализа данных		18	–	22

№	Тема самостоятельной работы	Учебно-методическое	Объем, ч		
		2020. - 46 с. - ISBN 978-5-7264-2193-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2196325 (дата обращения: 06.03.2025). – Режим доступа: по подписке. 4. Уорд, Б. Инновации SQL Server 2019. Использование технологий больших данных и машинного обучения / Боб Уорд ; пер. с англ. Н. Б. Желновой. - Москва : ДМК Пресс, 2020. - 408 с. - ISBN 978-5-97060-595-0. - Текст : электронный. - URL: https://znanium.com/catalog/product/1225362 (дата обращения: 06.03.2025). – Режим доступа: по подписке.			
Всего			144	–	172

4.6.5. Другие виды самостоятельной работы студентов

Не предусмотрены.

4.7. Перечень тем и видов занятий, проводимых в интерактивной форме

№ п/п	Форма занятия	Тема занятия	Интерактивный метод	Объем, ч
1.	Лекция	Введение в большие данные	Интерактивная лекция	2

5. Фонд оценочных средств для проведения промежуточной аттестации

Полное описание фонда оценочных средств текущей и промежуточной аттестации обучающихся с перечнем компетенций, описанием показателей и критериев оценивания компетенций, шкал оценивания, типовые контрольные задания и методические материалы представлены в Приложении 3 к настоящей программе.

6. Учебно-методическое обеспечение дисциплины

6.1. Рекомендуемая литература

6.1.1. Основная литература

№ п/п	Автор, название, место издания, изд-во, год издания	Кол-во экз. в библиот.
1.	Параскевов, А. В. Большие данные : учебное пособие / А. В. Параскевов, А. Э. Сергеев. - Москва ; Вологда : Инфра-Инженерия, 2024. - 148 с. - ISBN 978-5-9729-2120-1. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2169699 (дата обращения: 06.03.2025). –	Электронный ресурс

	Режим доступа: по подписке.	
2.	Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. - Москва : Изд. Дом МИСиС, 2016. - 52 с. - ISBN 978-5-87623-969-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/1232190 (дата обращения: 06.03.2025). – Режим доступа: по подписке.	Электронный ресурс
3.	Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов ; Министерство науки и высшего образования Российской Федерации, Национальный исследовательский Московский государственный строительный университет, кафедра информационных систем, технологий и автоматизации в строительстве. - Москва : Издательство МИСИ - МГСУ, 2020. - 46 с. - ISBN 978-5-7264-2193-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2196325 (дата обращения: 06.03.2025). – Режим доступа: по подписке.	Электронный ресурс
4.	Уорд, Б. Инновации SQL Server 2019. Использование технологий больших данных и машинного обучения / Боб Уорд ; пер. с англ. Н. Б. Желновой. - Москва : ДМК Пресс, 2020. - 408 с. - ISBN 978-5-97060-595-0. - Текст : электронный. - URL: https://znanium.com/catalog/product/1225362 (дата обращения: 06.03.2025). – Режим доступа: по подписке.	Электронный ресурс

6.1.2. Дополнительная литература

№ п/п	Автор, название, место издания, изд-во, год издания, количество страниц
1.	Лесковец, Ю. Анализ больших наборов данных : практическое руководство / Д. Дж. Ульман, Ю. Лесковец, А. Раджараман ; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2023. - 500 с. - ISBN 978-5-89818-304-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102592 (дата обращения: 06.03.2025). – Режим доступа: по подписке.
2.	Келлехер, Д. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни ; пер. с англ. - Москва : Альпина Паблишер, 2020. - 222 с. - ISBN 978-5-9614-3170-4. - Текст : электронный. - URL: https://znanium.ru/catalog/product/1221800 (дата обращения: 06.03.2025). – Режим доступа: по подписке.

6.1.3. Периодические издания

Не предусмотрены.

6.1.4. Методические указания для обучающихся по освоению дисциплины

Методические указания находятся в стадии разработки

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), необходимых для освоения дисциплины

№ п/п	Название интернет-ресурса, адрес и режим доступа
1.	Википедия – свободная энциклопедия. [Электронный ресурс]. URL: https://ru.wikipedia.org/ (дата обращения: 06.03.2025)
2.	Научная электронная библиотека «e-Library». [Электронный ресурс]. URL: https://elibrary.ru/ (дата обращения: 06.03.2025).
3.	Электронно-библиотечная система «Znanium» [Электронный ресурс]. URL: https://znanium.ru/ (дата обращения: 06.03.2025).

6.3. Средства обеспечения освоения дисциплины

6.3.1. Компьютерные обучающие и контролирующие программы

№ п/п	Вид учебного занятия	Наименование программного обеспечения	Функция программного обеспечения		
			контроль	моделирующая	обучающая
1	Лекционные, практические занятия, самостоятельная работа	http://moodle.lgau.ru	+	+	+

6.3.2. Аудио- и видеопособия

Не предусмотрены.

6.3.3. Компьютерные презентации учебных курсов

Не предусмотрены.

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

№ п/п	Наименование оборудованных учебных кабинетов, объектов для проведения занятий	Перечень основного оборудования, приборов и материалов
1.	Г-112 – компьютерный класс	Компьютеры – 7 шт., стол 1 тумб. – 1 шт., доска для тех. пок. – 1 шт., стул ученич. – 19 шт., стол компьют. – 7 шт., скам. аудит. – 2 шт., стол аудиторный – 7 шт.

8. Междисциплинарные связи

Протокол согласования рабочей программы с другими дисциплинами

Наименование дисциплины, с которой проводилось согласование	Кафедра, с которой проводилось согласование	Предложения об изменениях в рабочей программе. Заключение об итогах согласования
«Объектно-ориентированный анализ»	Кафедра информационных технологий, математики и физики	Согласовано
« OLAP&OLTP системы обработки и анализа данных»	Кафедра информационных технологий, математики и физики	Согласовано

Лист изменений рабочей программы

Номер изменения	Номер протокола заседания кафедры и дата	Страницы с изменениями	Перечень откоррек- тированных пунктов	Подпись заве- дующего кафедрой

Лист периодических проверок рабочей программы

Должностное лицо, проводившее проверку Ф.И.О., должность,	Дата	Потребность в корректировке	Перечень пунктов, стр., разделов, требующих изменений

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ УНИВЕРСИТЕТ ИМЕНИ К.Е. ВОРОШИЛОВА»**

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

учебной дисциплины «Основы работы с Большими данными»

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль): Бизнес-информатика

Уровень профессионального образования: бакалавриат

Год начала подготовки: 2025

Луганск, 2025

1. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ, СООТНЕСЕННЫХ С ИНДИКАТОРАМИ ДОСТИЖЕНИЯ КОМПЕТЕНЦИЙ, С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ В ПРОЦЕССЕ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Код контролируемой компетенции	Формулировка контролируемой компетенции	Индикаторы достижения компетенции	Этап (уровень) освоения компетенции	Планируемые результаты обучения	Наименование модулей и (или) разделов дисциплины	Наименование оценочного средства	
						Текущий контроль	Промежуточная аттестация
ПК-3	Способен разрабатывать модели бизнес-процессов заказчика с учетом требований к информационным системам	ПК-3.1. Составляет анкеты и интервью с предполагаемым и представителями заказчика и будущими пользователями системы, анализирует исходную документацию, описывает автоматизируемые с помощью информационных систем бизнес-процессы, формирует требования к компонентам информационной системы	Первый этап (пороговый уровень)	Знать: основы информационной и библиографической культуры работы с большими данными	Тема 1. Введение в большие данные Тема 2. Жизненный цикл аналитики данных	Тесты закрытого типа	Зачет
			Второй этап (продвинутый уровень)	Уметь: применять методы, приемы и способы решения стандартных задач профессиональной деятельности при работе с массивами больших данных	Тема 3. Высокопроизводительные вычисления Тема 4. Масштабирование и многоуровневое хранение данных	Тесты открытого типа (вопросы для опроса)	Зачет
			Третий этап (высокий уровень)	Иметь навыки: использования информационно-коммуникационных технологий, с учетом основных требований информационной безопасности, для сбора, обработки и анализа больших данных	Тема 5. Визуализация данных и результатов анализа Тема 6. Классификация задач анализа данных Тема 7. Сложные методы аналитики Тема 8. Основы построения нейронных сетей для анализа данных	Практические задания	Зачет
		ПК-3.2. Знает и применяет инструменты,	Первый этап (пороговый уровень)	Знать: основные понятия термина Большие Данные;		Тесты закрытого типа	Зачет

		методики описания и моделирования бизнес-процессов, осуществляет разработку моделей		основные понятия Data Mining; основные способы хранения и анализа данных; методы извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети).			
			Второй этап (продвинутый уровень)	Уметь: анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия		Тесты открытого типа (вопросы для опроса)	Зачет
			Третий этап (высокий уровень)	Иметь навыки: хранения и анализа данных; навыками извлечения знаний из данных		Практические задания	Зачет

				(кластеризация, классификация, прогнозирование, регрессия, нейронные сети)			
--	--	--	--	--	--	--	--

2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЯ НА РАЗЛИЧНЫХ ЭТАПАХ ИХ ФОРМИРОВАНИЯ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ

№ п/п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде	Критерии оценивания	Шкала оценивания
1.	Тест	Система стандартизированных заданий, позволяющая измерить уровень знаний.	Тестовые задания	В тесте выполнено 90-100% заданий	Оценка «Отлично» (5)
				В тесте выполнено более 75-89% заданий	Оценка «Хорошо» (4)
				В тесте выполнено 60-74% заданий	Оценка «Удовлетворительно» (3)
				В тесте выполнено менее 60% заданий	Оценка «Неудовлетворительно» (2)
				Большая часть определений не представлена, либо представлена с грубыми ошибками.	Оценка «Неудовлетворительно» (2)
2.	Опрос	Форма работы, которая позволяет оценить кругозор, умение логически построить ответ, умение продемонстрировать монологическую речь и иные коммуникативные навыки. Устный опрос обладает большими возможностями воспитательного воздействия, создавая условия для неформального общения.	Вопросы к опросу	Продемонстрированы предполагаемые ответы; правильно использован алгоритм обоснований во время рассуждений; есть логика рассуждений.	Оценка «Отлично» (5)
				Продемонстрированы предполагаемые ответы; есть логика рассуждений, но неточно использован алгоритм обоснований во время рассуждений и не все ответы полные.	Оценка «Хорошо» (4)
				Продемонстрированы предполагаемые ответы, но неправильно использован алгоритм обоснований во время рассуждений; отсутствует логика рассуждений; ответы не полные.	Оценка «Удовлетворительно» (3)
				Ответы не представлены.	Оценка «Неудовлетворительно» (2)
3.	Практические задания	Направлено на овладение методами и методиками изучаемой дисциплины. Для решения предлагается решить конкретное задание (ситуацию) без применения математических	Практические задания	Продемонстрировано свободное владение профессионально-понятийным аппаратом, владение методами и методиками дисциплины. Показаны способности самостоятельного мышления, творческой активности. Задание выполнено в полном объеме.	Оценка «Отлично» (5)

№ п/ п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде	Критерии оценивания	Шкала оценивания
		расчетов.		Продemonстрировано владение профессионально-понятийным аппаратом, при применении методов и методик дисциплины незначительные неточности, показаны способности самостоятельного мышления, творческой активности. Задание выполнено в полном объеме, но с некоторыми неточностями.	Оценка «Хорошо» (4)
				Продemonстрировано владение профессионально-понятийным аппаратом на низком уровне; допускаются ошибки при применении методов и методик дисциплины. Задание выполнено не полностью.	Оценка «Удовлетворительно» (3)
				Не продемонстрировано владение профессионально-понятийным аппаратом, методами и методиками дисциплины. Задание не выполнено.	Оценка «Неудовлетворительно» (2)
4.1	Экзамен	Контрольное мероприятие, которое проводится по окончании изучения дисциплины.	Вопросы к зачету	Показано знание теории вопроса, понятийно-терминологического аппарата дисциплины; умение анализировать проблему, содержательно и стилистически грамотно излагать суть вопроса; глубоко понимать материал; владение аналитическим способом изложения вопроса, научных идей; навыками аргументации и анализа фактов, событий, явлений, процессов. Выставляется обучающемуся, полно, подробно и грамотно	Оценка «Отлично» (5)

№ п/ п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде	Критерии оценивания	Шкала оценивания
				ответившему на вопросы билета и вопросы экзаменатора.	
				Показано знание основных теоретических положений вопроса; умение анализировать явления, факты, действия в рамках вопроса; содержательно и стилистически грамотно излагать суть вопроса, но имеет место недостаточная полнота ответов по излагаемому вопросу. Продemonстрировано владение аналитическим способом изложения вопроса и навыками аргументации. Выставляется обучающемуся, полностью ответившему на вопросы билета и вопросы экзаменатора, но допустив при ответах незначительные ошибки, указывающие на наличие несистемности и пробелов в знаниях.	Оценка «Хорошо» (4)
				Показано знание теории вопроса фрагментарно (неполнота изложения информации; оперирование понятиями на бытовом уровне); умение выделить главное, сформулировать выводы, показать связь в построении ответа не продемонстрировано. Владение аналитическим способом изложения вопроса и владение навыками аргументации не продемонстрировано. Обучающийся допустил несущественные ошибки при ответах на вопросы билетов и вопросы экзаменатора.	Оценка «Удовлетворительно» (3)
				Знание понятийного аппарата, теории вопроса, не продемонстрировано; умение анализировать учебный материал не продемонстрировано; владение аналитическим	Оценка «Неудовлетворительно» (2)

№ п/ п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде	Критерии оценивания	Шкала оценивания
				способом изложения вопроса и владение навыками аргументации не продемонстрировано. Обучающийся не ответил на один или два вопроса билета и дополнительные вопросы экзаменатора.	
4.2	Экзамен	Контрольное мероприятие, которое проводится по окончании изучения дисциплины.	Тестовые задания к экзамену	90-100% правильных ответов	Оценка «Отлично» (5)
				70-80% правильных ответов	Оценка «Хорошо» (4)
				60% правильных ответов	Оценка «Удовлетвори тельно» (3)
				10-50% правильных ответов	Оценка «Неудовлетвор ительно» (2)

3. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ПРОЦЕССЕ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Оценочные средства для проведения текущего контроля

Текущий контроль осуществляется преподавателем дисциплины при проведении занятий в форме тестовых заданий, устного опроса и практических заданий.

ПК-3. Способен разрабатывать модели бизнес-процессов заказчика с учетом требований к информационным системам.

ПК-3.1. Составляет анкеты и программы интервью с предполагаемыми представителями заказчика и будущими пользователями системы, анализирует исходную документацию, описывает автоматизируемые с помощью информационных систем бизнес-процессы, формирует требования к компонентам информационной системы.

Первый этап (пороговой уровень) – показывает сформированность показателя компетенции «знать»: основы информационной и библиографической культуры работы с большими данными.

Тестовые задания закрытого типа

1. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является: (выберите один вариант ответа)

- а) адаптированность
- б) не требуется дописывать программный код
- в) трудоемкость подготовки данных
- г) нет возможности подстроить программное обеспечение под нужды бизнеса

2. Постановка бизнес-задачи – это этап, который: (выберите один вариант ответа)

- а) формулирует конкретные бизнес-задачи, и они уже не могут быть изменены
- б) формулирует конкретные бизнес-задачи, и они могут быть изменены в ходе прохождения именно этого цикла
- в) формулирует конкретные бизнес-задачи, и они не могут быть изменены в ходе прохождения именно этого цикла
- г) формулирует цели на перспективу

3. SQL – это... (выберите один вариант ответа)

- а) реляционная база данных
- б) язык неструктурированных запросов
- в) язык структурированных запросов
- г) средство для создания пайплайнов

4. Hadoop – это...(выберите один вариант ответа)

- а) набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах
- б) распределённая СУБД, позволяющая обрабатывать большие данные
- в) язык выполнения заданий в парадигме MapReduce

г) распределённая файловая система, предназначенная для хранения файлов большого объёма

5. Отметьте верное понимание Variety в контексте характеристик Big Data...(выберите один вариант ответа)

- а) высокая скорость генерирования данных
- б) разные типы данных в колонках таблиц реляционных СУБД
- в) разнообразие отраслей, являющихся источниками данных
- г) разнообразие типов данных, включающих в себя структурированные, частично структурированные и неструктурированные

Ключи

1.	а
2.	в
3.	в
4.	а
5.	г

6. Прочитайте текст и установите последовательность.

Установите последовательность этапов процесса Data-Mining между ними соответствие.

Этапы процесса Data-Mining
а) Понимание бизнеса (Business understanding)
б) Подготовка данных (Data preparation)
в) Оценка (Evaluation)
г) Понимание и сбор данных (Data understanding and collection)
д) Моделирование (Modeling)
е) Развёртывание (Deployment)

Ключ

а, г, б, д, в, е

Второй этап (продвинутый уровень) – показывает сформированность показателя компетенции «уметь»: применять методы, приемы и способы решения стандартных задач профессиональной деятельности при работе с массивами больших данных.

Задания открытого типа (вопросы для опроса):

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Кто и в каком году впервые ввел термин «Big Data»?
4. Какие главные характеристики Big Data?
5. Какие понятия содержит в себе принцип трех "V"?

Ключи

1.	Big Data — это крупные массивы разнообразной информации и стек специальных технологий для работы с ней. Термин применяется к таким объемам данных, с которыми пользовательский компьютер и офисные программы не справятся.
2.	Главной целью обработки Big Data является быстрое и эффективное использование всех видов информации в условиях непрерывного изменения и прироста в больших объемах.

3.	Термин Big Data или «Большие Данные» появился в 2008 году. Журнал Nature посвятил спецвыпуск взрывному росту объемов информации, употребив это словосочетание. Считается, что именно благодаря редактору этого издания Клиффорду Линчу понятие вошло в обиход.
4.	Существует шесть основных характеристик Больших данных: скорость накопления данных, объем, разнообразие, достоверность, изменчивость и ценность.
5.	Характеристики определяющие так называемый принцип «Трёх V»(Volume, Velocity, Variety): объем (объем данных), скорость (скорость накопления), многообразие (информация различных форматов).

Третий этап (высокий уровень) – показывает сформированность показателя компетенции «иметь навыки»: использования информационно-коммуникационных технологий, с учетом основных требований информационной безопасности, для сбора, обработки и анализа больших данных.

Практические задания:

Задание 1. В ящике 10 одинаковых деталей, пронумерованных номерами от 1 до 10. Из ящика наугад вынимают шесть деталей. Найти вероятность того, что среди отобраанных деталей будет деталь номер 3?

Задание 2. Игральный кубик бросают два раза. Найти вероятность того, что сумма очков будет больше 10?

Задание 3. Круг вращения волчка разделен на 24 одинаковых сектора, которые пронумерованы от 1 до 24. Найти вероятность того, что метка волчка укажет на сектор с номером, кратным трем?

Задание 4. Изделие собирают из комплектующих, которые поступают с трех заводов. Доли комплектующих в изделии от каждого завода, соответственно, равны 0,5; 0,3; 0,2. Доля бракованных комплектующих среди поступивших, равна (по заводам) 0,01; 0,005; 0,015. Найти вероятность того, что собранное изделие будет бракованным, если в годном изделии бракованные изделия не допускаются.

Задание 5. Программист при написании программы делает ошибки. Из них синтаксические ошибки составляют долю 0,3, семантические ошибки составляют долю 0,55 и семиотические ошибки – 0,15. Компилятор не обнаруживает синтаксические ошибки с вероятностью 0,05, семантические ошибки с вероятностью 0,8 и семиотические ошибки с вероятностью 0,9.

Найти вероятность того, что написанная программистом программа после компиляции будет содержать ошибки.

Ключи

1.	0,6
2.	1/12
3.	1/3
4.	0,095
5.	0,59

ПК-3.2. Знает и применяет инструменты, методики описания и моделирования бизнес-процессов, осуществляет разработку моделей.

Первый этап (пороговой уровень) – показывает сформированность показателя компетенции «знать»: основные понятия термина Большие Данные; основные понятия Data Mining; основные способы хранения и анализа данных; методы извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети).

Тестовые задания закрытого типа

1. Дайте определение Big Data (выберите один вариант ответа)

- а) комплексный набор инструментов обработки структурированных данных колоссальных объемов
- б) комплексный набор подходов, инструментов и методов обработки структурированных и неструктурированных данных колоссальных объемов
- в) комплексный набор методов обработки неструктурированных данных
- г) комплексный набор методов обработки структурированных данных колоссальных объемов

2. Как происходит обучение нейронной сети? (выберите один вариант ответа)

- а) эксперты настраивают нейронную сеть
- б) сеть запускается на обучающем множестве, и незадействованные нейроны выкидываются
- в) сеть запускается на обучающем множестве, и подстраиваются весовые значения
- г) сеть запускается на обучающем множестве, и добавляются или убираются соединения между нейронами

3. Эффективно применять Data Mining с целью защиты от...(выберите один вариант ответа)

- а) мошенников
- б) инфицирующих ботов
- в) инфицированных клиентов
- г) инсайдеров

4. Примером применения Big Data не может быть...(выберите один вариант ответа)

- а) зондирование Земли из космоса
- б) родительский контроль
- в) хранение данных клиентов в соцсетях
- г) доступ к крупнейшим библиотекам

5. В краудсорсинге вклад в проект каждого участника осуществляется...(выберите один вариант ответа)

- а) лишь финансами
- б) лишь затратами времени
- в) ресурсами – компьютер, время и т.д.
- г) привлечением аутсорсинга

Ключи

1.	б
2.	в
3.	а
4.	б
5.	в

6. Прочитайте текст и установите соответствие

В таблице приведены понятия по теме обработки Больших данных и их определение. Установите между ними соответствие.

Определение термина	Термины
1. Процесс создания новых признаков из имеющихся данных, чтобы улучшить работу модели машинного	а) Кластеризация данных

обучения	
2. Способ наглядного представления результатов обработки данных с помощью графиков, диаграмм, сводных таблиц, блок-схем, инфографики и тепловых карт.	б) Feature Engineering
3. Процесс систематической организации необработанных данных в группы или категории на основе общих характеристик или атрибутов.	в) Визуализация данных
4. Разбиение множества объектов на подмножества по заданному критерию	г) Классификация данных
5. Процесс объединения и обобщения данных из разрозненных источников в единый набор данных.	д) Агрегация данных
	е) Фильтрация данных

Запишите в таблицу выбранные буквы под соответствующими цифрами

1	2	3	4	5
б	в	г	а	д

Второй этап (продвинутый уровень) – показывает сформированность показателя компетенции «уметь»: анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия.

Задания открытого типа (вопросы для опроса):

1. Перечислите стадии Data Mining.
2. Опишите технологию MapReduce?
3. Что означает термин NoSQL?
4. Основные принципы потоковой обработки данных.
5. Приведите примеры использования потоковой обработки событий.

Ключи

1.	1. Свободный поиск – выявление скрытых закономерностей 2. Валидация – проверка достоверности выявленных закономерностей 3. Прогностическое моделирование – предсказание неизвестных значений, на основе полученных на первой стадии закономерностей 4. Анализ исключений – выявление аномалий найденных закономерностей
2.	MapReduce — модель распределённых вычислений, представленная компанией Google, используемая для параллельных вычислений над очень большими (до нескольких петабайт) наборами данных в компьютерных кластерах. Суть MapReduce состоит в разделении информационного массива на части, параллельной обработки каждой части на отдельном узле и финального объединения всех результатов.
3.	Термин NoSQL применяется к хранилищам данных, которые не используют язык запросов SQL. Вместо этого они запрашивают данные с помощью других языков программирования и конструкций.
4.	– Постоянный поток данных. Data streaming предполагает обработку непрерывного потока данных, который может быть бесконечным или ограниченным временем. – Низкая задержка. Одна из ключевых особенностей потоковой обработки данных — это возможность обрабатывать данные с минимальной задержкой, порой даже в миллисекундах. – Учёт временной последовательности и порядка происхождения событий при

	обработке, даже если они достигли потокового процессора со смещением из-за сетевых задержек.
5.	– Финансовый сектор. Банки и финансовые организации используют потоковую обработку данных для мониторинга транзакций, оценки рисков и выявления мошенничества в режиме реального времени. – Транспорт и логистика. Компании используют потоковую обработку данных для отслеживания движения транспорта, оптимизации маршрутов и прогнозирования времени доставки грузов. – Рекомендательные системы. Потоковая обработка данных позволяет онлайн-магазинам анализировать пользовательские данные и предлагать рекомендации в режиме реального времени, увеличивая продажи и улучшая пользовательский опыт.

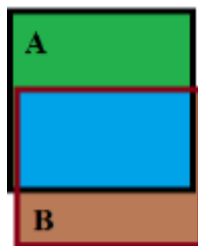
Третий этап (высокий уровень) – показывает сформированность показателя компетенции «иметь навыки»: терминологией Больших данных терминологией Data Mining; основными способами хранения и анализа данных; навыками извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети).

Практические задания:

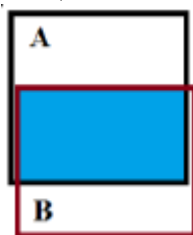
Задание 1. Данные, которые используются в программе, могут быть представлены постоянными значениями (константами) и изменяемыми, которые хранятся в переменных. Переменные обозначаются именами, которые могут содержать латинские буквы, цифры, знака подчёркивания и должны начинаться с буквы.

Если задано $a = I$, то это значит...

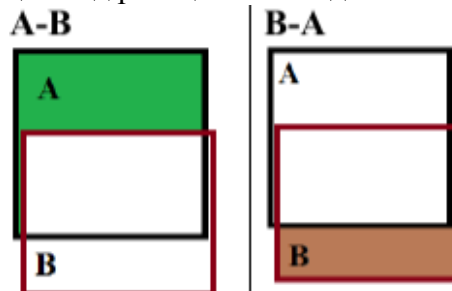
Задание 2. Какая операция над реляционными данными представлена?



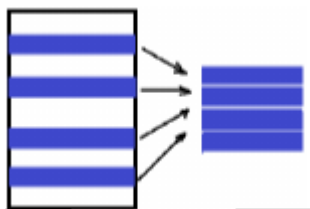
Задание 3. Какая операция над реляционными данными представлена?



Задание 4. Какая операция над реляционными данными представлена?



Задание 5. Какая операция над реляционными данными представлена?



Ключи

1.	переменная a будет иметь целый тип
2.	объединение
3.	пересечение
4.	вычитание (разность)
5.	прямое произведение

Оценочные средства для проведения промежуточной аттестации

Промежуточная аттестация проводится в виде экзамена.

Вопросы для экзамена

1. Понятие «Big Data» в информационных технологиях.
2. Основная цель обработки Big Data.
3. Главные характеристики Big Data.
4. Основные понятия принципа трех "V".
5. Квази-структурированные данные.
6. Возможные этапы работы с большими данными.
7. Обзор подходов к работе с данными: от языка простых запросов до методов анализа больших данных.
8. Интеллектуальный анализ данных: краткий обзор подходов.
9. Генетические алгоритмы.
10. Деревья принятия решений.
11. Визуализация больших данных.
12. Специфика хранения и обработки больших данных.
13. Пакет утилит, библиотек и фреймворков Apache Hadoop.
14. Преимущества решений на базе Hadoop.
15. Технология MapReduce.
16. Достоинства и недостатки MapReduce.
17. Документо-ориентированные хранилища.
18. Файловая система HDFS, принцип работы.
19. Функции NameNode в HDFS.
20. Репликация данных в Hadoop.
21. Компоненты узла Slave в Hadoop.
22. Компоненты узла Master в Hadoop.
23. Автономный режим Hadoop. Цели использования.
24. Оценка качества обучения модели.
25. Особенности NoSQL БД и отличия их от реляционных БД. Типы NoSQL баз данных.
26. Преимущества и недостатки NoSQL.
27. Ассоциативные правила.
28. Характеристики ассоциативных правил.

29. Методы поиска ассоциативных правил.
30. Анализ больших данных и традиционный анализа. Основные характеристики и отличительные черты.
31. Основные типы Data Mining
32. Категории Web Mining.
33. Основная задача Web Content Mining.
34. Основные задачи интеллектуального анализа текстов.
35. Модель данных HBase. Понятия строк, колонок, семейства колонок. Свойства семейств колонок. Временные метки.
36. Организация обмена данными между Spark и HBase.
37. Хранение данных в HBase. Файлы HFile, их организация.
38. Типы кластерных структур.
39. Типы визуализации. Требования предъявляются к визуализации.
40. MongoDB. Характеристика и способ хранения данных.

4. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРЫ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ

Для выполнения практических заданий студенту необходимы: ручка, листы для черновых подсчетов, калькулятор.

Текущий контроль

Тестирование для проведения текущего контроля проводится в виде тестов или системы дистанционного обучения Moodle.

На тестирование отводится 20 минут. Каждый вариант тестовых заданий включает 10 вопросов. Количество возможных вариантов ответов – 4. Студенту необходимо выбрать один правильный ответ. За каждый правильный ответ на вопрос присваивается 10 баллов. Шкала перевода: 9-10 правильных ответов – оценка «отлично» (5), 7-8 правильных ответов – оценка «хорошо» (4), 6 правильных ответов – оценка «удовлетворительно» (3), 1-5 правильных ответов – оценка «не удовлетворительно» (2).

Опрос как средство текущего контроля проводится в форме устных ответов на вопросы. Студент отвечает на поставленный вопрос сразу, время на подготовку к ответу не предоставляется.

Практические задания как средство текущего контроля проводятся в письменной форме. Студенту выдается задание и предоставляется 10 минут для подготовки к ответу.

Промежуточная аттестация

Экзамен проводится в устной форме. Из экзаменационных вопросов составляется 20 экзаменационных билетов. Каждый билет состоит из двух вопросов. Комплект экзаменационных билетов представлен в учебно-методическом комплексе дисциплины.

На подготовку к ответу студенту предоставляется 20 минут.