

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Дзержинский филиал ННГУ

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Технологии обработки больших данных

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

09.03.03 - Прикладная информатика

Направленность образовательной программы

ИТ-сервисы и технологии обработки данных в экономике и финансах

Форма обучения

очная, очно-заочная

г. Дзержинск

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.10 Технологии обработки больших данных относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	<p>УК-1.1: Демонстрирует знание принципов сбора, отбора и обобщения информации, базирующихся на системном подходе</p> <p>УК-1.2: Демонстрирует умение соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности</p> <p>УК-1.3: Демонстрирует наличие практического опыта работы с информационными источниками, опыта научного поиска и представления научных результатов</p>	<p>УК-1.1:</p> <p>Знать: понятие большие данные, методы сбора и обобщения информации.</p> <p>Уметь: обрабатывать и управлять большими объемами постоянно обновляющейся информации.</p> <p>Владеть: навыками сбора и обработки больших данных.</p> <p>УК-1.2:</p> <p>Знать: виды обработки больших данных: синхронная и асинхронная, пакетная и поточная.</p> <p>Уметь: решать стандартные задачи профессиональной деятельности на основе теории вероятностей и математической статистики.</p> <p>Владеть: навыками применения методов и средств обеспечения интеллектуального анализа данных.</p> <p>УК-1.3:</p> <p>Знать: технологии работы с информационными источниками.</p> <p>Уметь: работать с информационными источниками.</p> <p>Владеть: практическим</p>	<p>Отчет по лабораторным работам</p> <p>Собеседование</p> <p>Тест</p>	<p>Зачёт:</p> <p>Контрольные вопросы</p>

		<p>опытом работы с информационными источниками, навыками представления результатов обработки данных</p>		
<p>ПК-9: Способен моделировать прикладные (бизнес) процессы и объекты предметной области</p>	<p>ПК-9.1: Демонстрирует знание методических основ моделирования процессов и объектов предметной области</p> <p>ПК-9.2: Демонстрирует умение применения знаний к моделированию прикладных процессов и объектов предметной области при разработке программного обеспечения ИС</p> <p>ПК-9.3: Имеет практический опыт моделирования процессов и объектов на примере конкретной предметной области</p>	<p>ПК-9.1:</p> <p>Знать: методики и алгоритмы анализа и обработки больших данных с использованием современных языков программирования высокого уровня</p> <p>Уметь: применять методы математического анализа и моделирования к анализу больших данных.</p> <p>Владеть: приемами и алгоритмами обработки данных, и построения моделей восстановления данных.</p> <p>ПК-9.2:</p> <p>Знать: математические основы построения моделей анализа и обработки данных, способы извлечения дополнительной информации из массив данных.</p> <p>Уметь: обрабатывать данные, выявлять скрытые зависимости и закономерности в больших данных, извлеченных непосредственно или полученных в результате эксперимента.</p> <p>Владеть: навыками разработки моделей машинного обучения.</p> <p>ПК-9.3:</p> <p>Знать: современные инструменты, пакеты программ, и технологии обработки и анализа данных.</p> <p>Уметь: настраивать программное обеспечение, обеспечивающее доступ к удаленным данным, в том числе с применением облачных технологий.</p> <p>Владеть: навыками и</p>	<p>Отчет по лабораторным работам</p> <p>Собеседование</p> <p>Тест</p>	<p>Зачёт:</p> <p>Контрольные вопросы</p>

		<p>средствами экспериментов с использованием современных средств работы с большими данными.</p>		
--	--	---	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная	очно-заочная
Общая трудоемкость, з.е.	3	3
Часов по учебному плану	108	108
в том числе		
аудиторные занятия (контактная работа):		
- занятия лекционного типа	16	8
- занятия семинарского типа (практические занятия / лабораторные работы)	32	10
- КСР	1	1
самостоятельная работа	59	89
Промежуточная аттестация	0 Зачёт	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)		в том числе							
			Контактная работа (работа во взаимодействии с преподавателем), часы из них				Самостоятельная работа обучающегося, часы			
	Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
Тема 1. Введение в предмет «Технологии обработки больших данных»	11	10	2	1			2	1	9	9
Тема 2. Хранилища данных (ХД) и OLAP-системы	16	19	2	1	4	2	6	3	10	16
Тема 3. Задачи и практическое применение Data Mining	22	20	4	2	8	2	12	4	10	16
Тема 4. Методы кластеризации данных	22	20	4	2	8	2	12	4	10	16
Тема 5. Основные задачи в проблеме распознавания образов	20	19	2	1	8	2	10	3	10	16
Тема 6. Работа со сложными системами отклонений	16	19	2	1	4	2	6	3	10	16
Аттестация	0	0								
КСР	1	1					1	1		

Итого	108	108	16	8	32	10	49	19	59	89
-------	-----	-----	----	---	----	----	----	----	----	----

Содержание разделов и тем дисциплины

Тема 1. Введение в предмет «Технологии обработки больших данных»

Основные понятия Data Mining и возникновение, перспективы, проблемы Data mining. Системы поддержки принятия решений (СППР). Задачи систем поддержки принятия решений. База данных – СППР. Неэффективность использования OLTP-систем для анализа данных.

Тема 2. Хранилища данных (ХД) и OLAP-системы

Понятие данных. Значение понятий объект и атрибут, выборка, зависимая и независимая переменная. Типы шкал. Концепция хранилища данных. Организация ХД. Очистка данных. Хранилища данных и анализ. Многомерная модель данных. Определение OLAP-систем. Концептуальное многомерное представление данных. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Тема 3. Задачи и практическое применение Data Mining.

Интеллектуальный анализ данных. Модели и методы DM. Процесс обнаружения знаний. Инструменты Data Mining. Поиск ассоциативных правил. Постановка задачи. Секвенциальный анализ. Разновидности поиска ассоциативных правил. Методы представления результатов. Алгоритмы поиска ассоциативных правил.

Тема 4. Методы кластеризации данных.

Постановка задачи кластеризации. Меры близости, основанные на расстояниях, используемых в алгоритмах кластеризации. Базовые алгоритмы кластеризации. Адаптивные методы кластеризации. Классификация и регрессия. Постановка задачи. Представление результатов. Методы построения правил классификации. Методы построения математических функций. Сущность прогнозирования. Роль и значение прогнозирования. Понятия временных рядов, его компоненты и параметры. Задача визуализация данных.

Тема 5. Основные задачи в проблеме распознавания образов.

Проблема распознавания образов. Классификация и характеристика основных задач распознавания образов. Формирование информативного признакового пространства. Основные понятия. Критерия информативности. Логико-эвристические методы формирования признакового пространства. Статистические методы формирования признакового пространства. Алгоритмы распознавания, основанные на вычислении оценок (АВО). Необходимые определения. Этапы задания АВО. Теоремы о вычислении оценок. Задачи, решаемые алгоритмами вычисления оценок.

Тема 6. Работа со сложными системами отклонений.

Мера важности объектов в сложных системах. Важность объекта в системе. Мера важности. Примеры введения меры важности объектов в сложных системах. Мера важности признаков в задаче распознавания образов. Вычислительные алгоритмы получения информационных весов и оценки их сложности. Некоторые применения мера важности признаков.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:

- электронный курс "Технологии обработки больших данных"

(<https://e-learning.unn.ru/course/view.php?id=11686>).

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Отчет по лабораторным работам) для оценки сформированности компетенции УК-1:

Задание 1

1. Проверить нулевую гипотезу о том, что заданное значение $\mu = 10$ является математическим ожиданием нормально распределённой случайной величины при 5%-м уровне значимости для двусторонней критической области, если в результате обработки выборки объёма $n = 10$ получено выборочное среднее $\bar{x} = 12$, а выборочное среднее квадратическое отклонение равно $S_1 = 1$.

2. При уровне значимости $\alpha = 0.05$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин X и Y на основе выборочных данных при альтернативной гипотезе :

$$\sigma_x^2 \neq \sigma_y^2$$

x_i	20	22	23	24	26	y_i	18	19	20	22	23
n_i	3	4	2	2	4	n_i	6	3	4	2	5

5.1.2 Типовые задания (оценочное средство - Отчет по лабораторным работам) для оценки сформированности компетенции ПК-9:

Задание 1

В таблице приведены данные по двум экономическим показателям

1. Рассчитайте параметры уравнений линейной, степенной, показательной и гиперболической парных регрессий.

2. Оцените тесноту связи с помощью показателей корреляции и детерминации.

3. Оцените качество уравнений с помощью средней ошибки аппроксимации.

4. Оцените статистическую надёжность результатов регрессионного моделирования с помощью F-критерия Фишера. По значениям характеристик, рассчитанных в пп. 3, 4 и данном пункте, выберите лучшее уравнение регрессии и дайте обоснование выбору.

5. Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 10 % от его среднего уровня.

6. Оцените полученные результаты.

Номер района	Потребительские расходы в расчёте на душу населения, тыс. руб., у	Среднемесячный доход на душу населения, тыс. руб., х
1	10,07	18,47
2	12,00	18,67
3	10,33	18,17
4	13,83	22,40
5	15,07	26,53
6	16,73	25,90
7	11,83	21,07
8	13,87	22,93
9	16,70	27,77
10	13,43	19,23
11	6,93	19,47
12	15,40	31,63
13	12,27	29,60
14	13,30	27,70
15	11,40	18,73
16	11,80	22,17
17	18,60	23,50

Задание 2

Администрация банка изучает динамику депозитов физических лиц за ряд лет (млрд. долл. В сопоставимых ценах). Исходные данные представлены в таблице.

Время, лет	1	2	3	4	5	6	7
Депозиты физических лиц, х	2	6	7	3	10	12	13

Известно, что $\sum x^2 = 511$

1. Постройте уравнение линейного тренда и дайте интерпретацию его параметров.
2. Определите коэффициент детерминации для линейного тренда.
3. Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2,5 млрд. долл. Подтверждается ли это полученными Вами результатами?

Задание 3

При уровне значимости 0,05 методом дисперсионного анализа проверьте нулевую гипотезу о влиянии фактора на качество объекта на основании пяти измерений для трёх уровней фактора Ф:

Номер измерения	Φ_1	Φ_2	Φ_3
1	12	10	20
2	16	8	26
3	15	7	28
4	17	5	24
5	14	9	27

Критерии оценивания (оценочное средство - Отчет по лабораторным работам)

Оценка	Критерии оценивания
превосходно	если выполнены 100% условий и требований, сформулированных в задании.

Оценка	Критерии оценивания
отлично	если выполнены 95% – 99% условий и требований, сформулированных в задании.
очень хорошо	если выполнены 94% условий и требований, сформулированных в задании.
хорошо	если выполнены 70% – 93% условий и требований, сформулированных в задании.
удовлетворительно	если выполнены 40% – 70% условий и требований, сформулированных в задании.
неудовлетворительно	если выполнены 20% - 40% условий и требований, сформулированных в задании.
плохо	если выполнены менее 20% условий и требований, сформулированных в задании.

5.1.3 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции УК-1:

Задача анализа данных. Технологии обработки данных. Отличия от других видов анализа данных.

Постановка и порядок решение задачи интеллектуального анализа данных на примере информационной модели торговой компании.

Закономерности в данных: неочевидные, объективные, полезные. Области применения интеллектуального анализа данных.

Подготовка данных для анализа. Обработка пропущенных и недостающих данных. Анализ исключений.

Виды и особенности шкал измерений данных.

Постановка и порядок решения задачи интеллектуального анализа данных на примере документной базы данных.

Особенности использования данных при интеллектуальном анализе данных. Данные, информация и знания.

Основные задачи интеллектуального анализа данных.

Стадии интеллектуального анализа данных. Начальные этапы. Вычисление и анализ показателей описательной статистики.

5.1.4 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции ПК-9:

Основные концепции баз данных. Модель данных. СУБД. Использование баз данных в интеллектуальном анализе данных.

Постановка и порядок решения задачи интеллектуального анализа данных на примере хранилища слабоструктурированных текстовых документов.

Введение в многомерные базы данных. Особенности базы данных для оперативной аналитической обработки данных. Порядок и особенности построения хранилищ данных.

Введение в OLAP. Типовая модель данных для OLAP. Особенности приложений для оперативной аналитической обработки данных.

Характеристики временных рядов, используемые в ходе интеллектуального анализа данных.

Построение и использование моделей для интеллектуального анализа данных.

Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.

Преимущественные особенности методов анализа данных: статистических, машинного обучения, оперативного анализа, интеллектуального анализа.

Решение задач кластеризация в ходе интеллектуального анализа данных: цель и основные методы.

Сущность и методы решения задачи прогнозирования при интеллектуальном анализе данных.

Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных в СППР.

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
превосходно	правильное понимание сущности вопроса, грамотность и логичность в изложении ответа, обосновании выводов; знание категориально-понятийного аппарата по теме; изложение материала в полном объеме; установление связи между изучаемым и ранее изученным материалом.
отлично	знание категориально-понятийного аппарата по теме; изложение материала в полном объеме; установление связи между изучаемым и ранее изученным материалом.
очень хорошо	правильное понимание сущности вопроса и логичность в изложении ответа; знание категориально-понятийного аппарата по теме; присутствуют недочеты, которые исправляются обучающимся самостоятельно или с помощью наводящих вопросов (уточнений) преподавателя.

Оценка	Критерии оценивания
хорошо	знание категориально-понятийного аппарата по теме; присутствуют недочеты, которые исправляются обучающимся самостоятельно или с помощью наводящих вопросов (уточнений) преподавателя.
удовлетворительно	правильное понимание сущности вопроса, но частичное изложение, непоследовательность материала, допущены неточности в определении понятий или формулировке правил; со стороны преподавателя требуется большое количество наводящих вопросов по проблеме; допущен ряд грубых ошибок.
неудовлетворительно	обучающийся не овладел основными знаниями и умениями в соответствии с требованиями программы.
плохо	обучающийся отказался от ответа на заданный преподавателем вопрос.

5.1.5 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции УК-1:

1. Для машинного обучения подходят данные

A) Любых форматов в цифровом виде

B) Числовые типа int

C) Бинарные

D) Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов

2. Укажите фактор, способствовавший появлению тренда больших данных

A) маркетинговые кампании крупных корпораций

B) снижение издержек на хранение данных

C) появление новых технологий обработки потоковых данных

D) выпуск баз данных с обработкой данных в памяти

3. Выберите верный ответ

A) большие данные – это обработка или хранение более 1 Тб информации

B) проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна

С) большие данные – это огромная PR-акция крупных вендоров и не более того

Д) большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект

4. Отметьте те из вариантов, в которых данные структурированы:

А) данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word

В) таблица с ежедневными показаниями температуры помещения за год в файле формата csv

С) текст педагогической поэмы А.С. Макаренко, представленный в формате PDF

Д) библиотека фильмов, представленных в формате mp4 на одном жестком диске

5. Какие из следующих средств разумно использовать для анализа данных, представленных единственным csv-файлом размера более 100Гб:

А) Hadoop

В) Data Warehouse

С) "Песочница"

Д) Python

5.1.6 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-9:

1. Недостатками подхода, при котором имеется хранилище данных (ХД) и витрины данных (ВД) для информации по разделам данной области, являются:

А) увеличение объема данных, хранимых в ВД

В) дополнительные затраты на разработку систем поддержки принятия решений с ХД и ВД

С) увеличение нагрузки на основное ХД

Д) избыточность (данные хранятся как в ХД, так и в ВД)

2. Данные в хранилище данных делятся на следующие категории:

А) детальные данные

В) агрегированные данные

С) метаданные

D) временные данные

3. Агрегированные данные в хранилище данных подразделяются на следующие типы:

A) аддитивные

B) транзитивные

C) полуаддитивные

D) неаддитивные

4. Метаданные описывают:

A) объекты предметной области, информация о которых хранится в хранилище данных

B) категории пользователей, использующих данные

C) местоположение серверов, рабочих станций и оперативные источники данных

D) системных администраторов

E) размещенные на серверах и рабочих станциях программных средств и распределение данных

5. Данные, поступающие из оперативных источников данных в хранилища данных, образуют следующие информационные потоки:

A) входной

B) виртуальных данных

C) метаданных

D) обобщения

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
превосходно	Процент выполнения тестовых заданий 100 %
отлично	Процент выполнения тестовых заданий 95 – 99 %
очень хорошо	Процент выполнения тестовых заданий 94%

Оценка	Критерии оценивания
хорошо	Процент выполнения тестовых заданий 75 – 93%
удовлетворительно	Процент выполнения тестовых заданий 50 – 74 %
неудовлетворительно	Процент выполнения тестовых заданий 20 - 50 %
плохо	Процент выполнения тестовых заданий < 20 %

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов

<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач
---------------	--	---	--	--	--	--	---

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции УК-1

Задача анализа данных. Технологии обработки данных. Отличия от других видов анализа данных.

Постановка и порядок решение задачи интеллектуального анализа данных на примере информационной модели торговой компании.

Закономерности в данных: неочевидные, объективные, полезные. Области применения интеллектуального анализа данных.

Подготовка данных для анализа. Обработка пропущенных и недостающих данных. Анализ исключений.

Виды и особенности шкал измерений данных.

Постановка и порядок решения задачи интеллектуального анализа данных на примере документной базы данных.

Особенности использования данных при интеллектуальном анализе данных. Данные, информация и знания.

Основные задачи интеллектуального анализа данных.

Стадии интеллектуального анализа данных. Начальные этапы. Вычисление и анализ показателей описательной статистики.

5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-9

Основные концепции баз данных. Модель данных. СУБД. Использование баз данных в интеллектуальном анализе данных.

Постановка и порядок решения задачи интеллектуального анализа данных на примере хранилища слабоструктурированных текстовых документов.

Введение в многомерные базы данных. Особенности базы данных для оперативной аналитической обработки данных. Порядок и особенности построения хранилищ данных.

Введение в OLAP. Типовая модель данных для OLAP. Особенности приложений для оперативной аналитической обработки данных.

Характеристики временных рядов, используемые в ходе интеллектуального анализа данных.

Построение и использование моделей для интеллектуального анализа данных.

Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.

Преимущественные особенности методов анализа данных: статистических, машинного обучения, оперативного анализа, интеллектуального анализа.

Решение задач кластеризация в ходе интеллектуального анализа данных: цель и основные методы.

Сущность и методы решения задачи прогнозирования при интеллектуальном анализе данных.

Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных в СППР.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»

Оценка	Критерии оценивания
не зачтено	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Макшанов А. В. Технологии интеллектуального анализа данных : учебное пособие / Макшанов А. В., Журавлев А. Е. - 2-е изд., стер. - Санкт-Петербург : Лань, 2022. - 212 с. - Книга из коллекции Лань - Информатика. - ISBN 978-5-8114-4493-9., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=798859&idb=0>.
2. Храмов А. Г. Методы и алгоритмы интеллектуального анализа данных : учебное пособие / Храмов А. Г. - Самара : СамГУ, 2019. - 176 с. - Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для обучающихся по основной образовательной программе высшего образования по направлению подготовки 01.04.02 Прикладная математика и информатика. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции СамГУ - Информатика. - ISBN 978-5-7883-1414-3., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=728573&idb=0>.

Дополнительная литература:

1. Нестеров С. А. Основы интеллектуального анализа данных. Лабораторный практикум : учебное пособие / Нестеров С. А. - Санкт-Петербург : Лань, 2020. - 40 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции Лань - Информатика. - ISBN 978-5-8114-4509-7., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=707838&idb=0>.
2. Станкевич Л. А. Интеллектуальные системы и технологии / Станкевич Л. А. - Москва : Юрайт, 2022. - 397 с. - (Высшее образование). - URL: <https://urait.ru/bcode/489694> (дата обращения: 05.01.2022). - ISBN 978-5-534-02126-4 : 1209.00. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=784455&idb=0>.
3. Железнов М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / Железнов М. М. - Москва : МИСИ – МГСУ, 2020. - 46 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции МИСИ – МГСУ - Информатика. - ISBN 978-5-7264-2193-3., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=746030&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. Электронная библиотека учебников [Электронный ресурс]. - Режим доступа: <http://studentam.net>
2. Российская государственная библиотека [Электронный ресурс]. - Режим доступа: <http://www.rsl.ru>

3. Научная электронная библиотека [Электронный ресурс]. - Режим доступа: <http://elibrary.ru/>
4. Программное обеспечение: менеджер пакетов Anaconda, Spyder (open-source IDE для Python).

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 09.03.03 - Прикладная информатика.

Автор(ы): Кечкина Наталия Игоревна, кандидат технических наук.

Заведующий кафедрой: Поляков Евгений Артурович, кандидат педагогических наук.

Программа одобрена на заседании методической комиссии от 22.12.2023, протокол № 17.