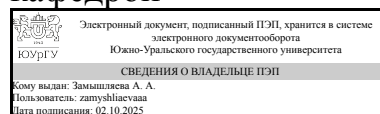


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П0.02 Интеллектуальный анализ данных
для направления 02.03.02 Фундаментальная информатика и информационные технологии

уровень Бакалавриат

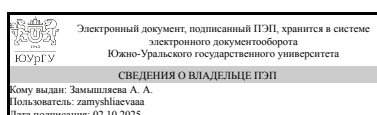
профиль подготовки Искусственный интеллект, глубокое обучение и анализ данных

форма обучения очная

кафедра-разработчик Центр ОП топ-уровня в сфере ИИ "ВиртУм"

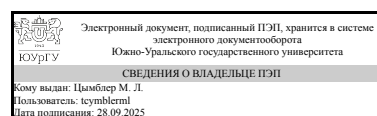
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 808

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



А. А. Замышляева

Разработчик программы,
д.физ.-мат.н., доц., профессор



М. Л. Цымблер

1. Цели и задачи дисциплины

Целью дисциплины является ознакомление студентов с основными методами и алгоритмами интеллектуального анализа данных (ИАД) — поиск шаблонов, классификация, кластеризация, поиск аномалий, а также формирование навыков идентификации базовых задач ИАД в реальных предметных областях и умений применять изученные методы и алгоритмы для их решения.

Краткое содержание дисциплины

Введение в дисциплину. Феномен Больших данных. Понятие интеллектуального анализа данных. Технологический цикл анализа данных. Базовые задачи интеллектуального анализа данных: поиск шаблонов, классификация, кластеризация, поиск аномалий. Поиск шаблонов. Понятия транзакции, частого набора, шаблона, поддержки, достоверности. Основные алгоритмы поиска частых наборов (Apriori, FP-Growth, Eclat). Выбор полезных шаблонов. Компактное представление частых наборов. Классификация. Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Наивная Байесовская классификация. Классификация по ближайшим соседям. Оценка качества классификации (метрики Accuracy, Precision, Recall, F1; AUC ROC; подготовка тестовой выборки: hold-out, random sampling, bootstrapping, cross-validation). Ансамблевая классификация (бэггинг, бустинг, случайный лес и др.). Кластеризация. Разделительная кластеризация (k-means, k-medoids и др.). Иерархическая кластеризация (Single linkage, Full linkage, Ward distance и др.). Плотностная кластеризация (DBSCAN). Нечеткая кластеризация (Fuzzy C-means). Оценка качества кластеризации (метод локтя, силуэтный коэффициент и др.). Поиск аномалий. Понятия аномалии (выброса), шума, новизны в данных. Виды аномалий: точечные, коллективные, контекстные. Статистические методы поиска аномалий. Поиск аномалий на основе расстояния. Поиск аномалий на основе плотности. Поиск аномалий с помощью кластеризации. Поиск аномалий на основе классификации.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-5 [ML-3] Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения	Знает: - [И-1, ПУ] стандартные библиотеки для обучения с учителем (ScikitLearn); базовые методы классификации с помощью деревьев решений Умеет: - [И-1, ПУ] обосновывать выбор конкретных алгоритмов и их параметров в зависимости от задачи и данных Имеет практический опыт: - [И-1, ПУ] использования метрик качества классификации (precision, recall, F1, AUC ROC и др.); комбинирования различных методов для комплексного анализа данных; интерпретации результатов в контексте предметной области
ПК-6 [ML-4] Способен применять методы обучения без учителя для анализа структуры	Знает: - [И-1, БУ] стандартные библиотеки для обучения без учителя (ScikitLearn); базовые

данных и выявления скрытых закономерностей	модели для кластерного анализа (неиерархическая кластеризация - k-средних dbSCAN иерархическая кластеризация) методы понижения размерности (PCA t-SNE) Умеет: - [И-2, ПУ] настраивать и применять алгоритмы обнаружения аномалий (статистические методы isolation forest one-class SVM) и ассоциативного анализа (Apriori, FP-Growth) с учётом структуры и особенностей реальных данных Имеет практический опыт: - [И-3, ПУ] использования метрик качества кластеризации (silhouette score adjusted rand index); комбинирования различных методов для комплексного анализа данных; интерпретации результатов в контексте предметной области
ПК-8 [BD-1] Способен осуществлять поиск, сбор, очистку и предварительный анализ данных	Умеет: - [И-2, ПУ] производить очистку зашумленных временных рядов и изображений

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Сбор, анализ и предобработка данных в машинном обучении	Анализ временных рядов, Производственная практика (технологическая, проектно-технологическая, стажировка) (6 семестр)

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
Сбор, анализ и предобработка данных в машинном обучении	Знает: -[И-1, БУ] основные числовые характеристики статистических данных и методы их нахождения, источники данных: типы и формы представления данных (структурированные, полуструктурированные, неструктурированные); методы статистической обработки и визуализации данных; методологии сбора данных; современные технологии хранения больших объемов данных, -[И-1, БУ] основные требования к наборам данных для решения задач машинного обучения, -[И-1, ПУ] опасности предвзятых данных Умеет: -[И-1, БУ] вычислять числовые характеристики статистических данных, применять методы визуализации данных, собирать и интегрировать разнородные наборы данных из разных источников, -[И-2, ПУ] подбирать инструментарий разметки данных под условия задачи, -[И-1, ПУ] проверять обучающую выборку на наличие различных искажений Имеет практический опыт: поиска и извлечения

	необходимых данных из различных источников; организации хранилищ данных и настройки инфраструктуры для хранения и быстрого доступа к данным, -[И-1, БУ] проверки данных на корректность, разметки данных, -[И-1, ПУ] анализа обучающей выборки на предмет репрезентативности, возможных искажений, скрытых предвзятостей
--	--

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 70,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		4
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	64	64
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0
Лабораторные работы (ЛР)	32	32
<i>Самостоятельная работа (СРС)</i>	37,75	37,75
Индивидуальное исследовательское задание по выбранной маломасштабной задаче (подзадаче) индустриального партнера	37,75	37.75
Консультации и промежуточная аттестация	6,25	6,25
Вид контроля (зачет, диф.зачет, экзамен)	-	зачет

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Введение в дисциплину	2	2	0	0
2	Поиск шаблонов	16	8	0	8
3	Классификация	16	8	0	8
4	Кластеризация	16	8	0	8
5	Поиск аномалий	14	6	0	8

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Феномен Больших данных. Понятие интеллектуального анализа данных. Технологический цикл анализа данных. Основные задачи анализа данных: поиск шаблонов, классификация, кластеризация, поиск аномалий.	2
2	2	Понятия транзакции, частого набора, шаблона, поддержки, достоверности. Основные алгоритмы поиска частых наборов: Apriori, Eclat, FP-Growth.	4

3	2	Выбор полезных шаблонов на основе мер support, confidence, lift и др. Компактное представление частых наборов: максимально частые и замкнутые наборы, иерархии наборов. Фрагментация и сэмплинг для поиска частых наборов	4
4	3	Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Меры оценки доли примесей в узле дерева решений: индекс Джини, энтропия; алгоритмы классификации ID3, C4.5, CART. Байесовская классификация. Классификация по ближайшим соседям. Оценка качества классификации: меры Accuracy, Precision, Recall, F1.	6
5	3	Ансамблевая классификация: бэггинг, бустинг, случайный лес.	2
6	4	Задачи кластеризации данных и подходы к ее решению. Разделительная кластеризация: алгоритмы k-means, k-medoids и др.	3
7	4	Иерархическая кластеризация: дендрограммы, агломеративный и дивизимный подход. Меры схожести кластеров: Single linkage, Complete linkage, Group average и др. Плотностная кластеризация: алгоритм DBSCAN. Нечеткая кластеризация: алгоритм Fuzzy C-Means.	3
8	4	Меры качества кластеризации: критерий Хопкинса, кросс-валидация, метод локтя, силуэтный коэффициент и др.	2
9	5	Понятия аномалии (выброса), шума, новизны в данных. Виды аномалий: точечные, глобальные, контекстные, смешанные. Статистические методы поиска аномалий: z-значимость, правило трех сигм, гистограммы. Поиск аномалий на основе расстояния. Поиск аномалий на основе плотности: метод вложенных циклов, метод решеток. Поиск аномалий с помощью разделительной и плотностной кластеризации. Поиск аномалий на основе классификации: метод One Class SVM, метод изолирующего леса.	6

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	2	Поиск частых наборов с помощью алгоритмов Apriori, ECLAT и FP-Growth.	4
2	2	Поиск шаблонов с помощью мер support, confidence, lift.	4
3	3	Байесовская классификация. Классификация с помощью дерева решений.	4
4	3	Ансамблевая классификация с помощью бэггинга. Ансамблевая классификация с помощью случайного леса. Ансамблевая классификация с помощью бустинга	4
4	4	Разделительная кластеризация с помощью алгоритмов k-Means и k-Medoids.	2
5	4	Плотностная кластеризация с помощью алгоритма DBSCAN.	2
6	4	Иерархическая кластеризация с помощью различных мер схожести кластеров.	2
7	4	Оценка качества кластеризации.	2
8	5	Поиск точечных аномалий с помощью z-значимости, правила трех сигм, гистограмм.	4
9	5	Поиск коллективных аномалий метода вложенных циклов, метода решеток, кластеризации.	4

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Индивидуальное исследовательское задание по выбранной маломасштабной задаче (подзадаче) индустриального партнера	Tan P.-N., Steinbach M., Karpatne A., Kumar V. Introduction to Data Mining. 2nd Edition. Pearson, 2019. 839 p. Chapter 10. Avoiding False Discoveries, p. 750-808.	4	37,75

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	4	Текущий контроль	Контрольный опрос по теме "Введение в дисциплину"	1	10	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Введение в дисциплину". Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос дает один балл. Время на прохождение теста – не менее 15 минут.	зачет
2	4	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск частых наборов"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и	зачет

					предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
3	4	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск ассоциативных правил" (выполняется на наборах данных из задач промышленных партнеров)	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора;	зачет

						– при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.	
4	4	Текущий контроль	Контрольный опрос по теме "Поиск шаблонов"	1	10	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Поиск шаблонов". Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос дает один балл. Время на прохождение теста – не менее 15 минут.	зачет
5	4	Текущий контроль	Практическое задание "Классификация. Байесовская классификация"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода	зачет

					содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
6	4	Текущий контроль	Практическое задание "Классификация. Деревья решений" (выполняется на наборах данных из задач промышленных партнеров)	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии	зачет

					с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
7	4	Текущий контроль	Практическое задание "Классификация. Бэггинг"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и	зачет

					описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
8	4	Текущий контроль	Практическое задание "Классификация. Случайный лес" (выполняется на наборах данных из задач индустриальных партнеров)	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и	зачет

					выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
9	4	Текущий контроль	Практическое задание "Классификация. Бустинг"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров	зачет

					соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
10	4	Текущий контроль	Контрольный опрос по теме "Классификация"	1	10	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Классификация". Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос дает один балл. Время на прохождение теста – не менее 15 минут.	зачет
11	4	Текущий контроль	Практическое задание "Кластеризация. Разделительная кластеризация"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом	зачет

					программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
12	4	Текущий контроль	Практическое задание "Кластеризация. Плотностная кластеризация" (выполняется на наборах данных из задач индустриальных партнеров)	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует	зачет

					поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
13	4	Текущий контроль	Практическое задание "Кластеризация. Иерархическая кластеризация" (выполняется на наборах данных из задач промышленных партнеров)	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче;	зачет

					– код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
14	4	Текущий контроль	Практическое задание "Кластеризация. Качество кластеризации"	3	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и	зачет

					подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
15	4	Текущий контроль	Контрольный опрос по теме "Кластеризация"	1	10	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Кластеризация". Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос дает один балл. Время на прохождение теста – не менее 15 минут.	зачет
16	4	Текущий контроль	Практическое задание "Поиск аномалий. Поиск точечных аномалий"	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2	зачет

						баллов: 0 - не выполнен, 1 - выполнен частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.	
17	4	Текущий контроль	Практическое задание "Поиск аномалий. Поиск коллективных аномалий" (выполняется на	3,5	10	Выполнение задания оценивается от 0 до 10 баллов. Максимальная оценка выставляется при полном выполнении каждого из следующих пяти критериев (критерий оценивается от 0 до 2 баллов: 0 - не выполнен, 1 - выполнен	зачет

			наборах данных из задач индустриальных партнеров)		частично, 2 - выполнен полностью): 1. Корректное решение поставленной задачи – представленный студентом программный код четко соответствует поставленной задаче; – код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно); – код компилируется без синтаксических ошибок и предупреждений компилятора; – при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты. 2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании. 3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы. 4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию. 5. Готовность отчета – студентом подготовлен отчет о выполнении задания, представляющий собой связный и структурированный документ со следующей информацией: формулировка задания; гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами; рисунки с результатами визуализации; пояснения, раскрывающие смысл полученных результатов.		
18	4	Текущий контроль	Контрольный опрос по теме "Поиск аномалий"	1	10	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Поиск аномалий". Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос дает	зачет

	ROC и др.); комбинирования различных методов для комплексного анализа данных; интерпретации результатов в контексте предметной области																				
ПК-6	Знает: - [И-1, БУ] стандартные библиотеки для обучения без учителя (ScikitLearn); базовые модели для кластерного анализа (неиерархическая кластеризация - k-средних dbscan иерархическая кластеризация) методы понижения размерности (PCA t-SNE)													+	+	+	+	+		+	
ПК-6	Умеет: - [И-2, ПУ] настраивать и применять алгоритмы обнаружения аномалий (статистические методы isolation forest one-class SVM) и ассоциативного анализа (Apriori, FP-Growth) с учётом структуры и особенностей реальных данных		+	+	+													+	+	+	+
ПК-6	Имеет практический опыт: - [И-3, ПУ] использования метрик качества кластеризации (silhouette score adjusted rand index); комбинирования различных методов для комплексного анализа данных; интерпретации результатов в контексте предметной области													+	+	+	+	+		+	
ПК-8	Умеет: - [И-2, ПУ] производить очистку зашумленных временных рядов и изображений																	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) *основная литература:*

Не предусмотрена

б) *дополнительная литература:*

Не предусмотрена

в) *отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:*

Не предусмотрены

г) *методические указания для студентов по освоению дисциплины:*

1. Dutt A., Ismail M.A., Herawan T., Abaker I., Hashem T. Partition-Based Clustering Algorithms Applied to Mixed Data for Educational Data Mining: A Survey From 1971 to 2024. IEEE Access. 2024. Vol. 12. P. 172923-172942. <https://doi.org/10.1109/ACCESS.2024.3496929>

2. Ki C., Sivakumar R., Mulerikkal J.P., Ayyappan B., Gupta M., Jan T. A comprehensive survey of machine learning and deep learning approaches for anomaly detection in high-performance computing systems. J. Supercomput. 2025. Vol. 81, no. 8). Article. 1032. <https://doi.org/10.1007/s11227-025-07503-4>

3. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.
4. Методические указания к практическим занятиям
5. Galbrun E. The minimum description length principle for pattern mining: a survey. Data Min. Knowl. Discov. 2022. Vo. 36, no. 5. P. 1679-1727. <https://doi.org/10.1007/s10618-022-00846-z>
6. Hassan D.O., Hassan B.A. A comprehensive systematic review of machine learning in the retail industry: classifications, limitations, opportunities, and challenges. Neural Comput. Appl. 2025. Vol. 37, np. 4. P. 2035-2070. <https://doi.org/10.1007/s00521-024-10869-w>

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Dutt A., Ismail M.A., Herawan T., Abaker I., Hashem T. Partition-Based Clustering Algorithms Applied to Mixed Data for Educational Data Mining: A Survey From 1971 to 2024. IEEE Access. 2024. Vol. 12. P. 172923-172942. <https://doi.org/10.1109/ACCESS.2024.3496929>
2. Ki C., Sivakumar R., Mulerikkal J.P., Ayyappan B., Gupta M., Jan T. A comprehensive survey of machine learning and deep learning approaches for anomaly detection in high-performance computing systems. J. Supercomput. 2025. Vol. 81, no. 8). Article. 1032. <https://doi.org/10.1007/s11227-025-07503-4>
3. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.
4. Galbrun E. The minimum description length principle for pattern mining: a survey. Data Min. Knowl. Discov. 2022. Vo. 36, no. 5. P. 1679-1727. <https://doi.org/10.1007/s10618-022-00846-z>
5. Hassan D.O., Hassan B.A. A comprehensive systematic review of machine learning in the retail industry: classifications, limitations, opportunities, and challenges. Neural Comput. Appl. 2025. Vol. 37, np. 4. P. 2035-2070. <https://doi.org/10.1007/s00521-024-10869-w>

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	eLIBRARY.RU	Алексеев Д.С., Щекочихин О.В. Технологии интеллектуального анализа данных. Учебное пособие. Кострома, 2020. 140 с. https://elibrary.ru/item.asp?id=43946965
2	Дополнительная литература	eLIBRARY.RU	Жаров А.Н., Минеичева И.Г. Анализ данных. Ярославль, 2020. 148 с. https://elibrary.ru/item.asp?id=43846458
3	Методические пособия для самостоятельной работы студента	Образовательная платформа Юрайт	Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с. https://urait.ru/bcode/432851

Перечень используемого программного обеспечения:

1. -Deductor Academic(01.09.2023)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Практические занятия и семинары		Персональный компьютер
Лекции		Проектор