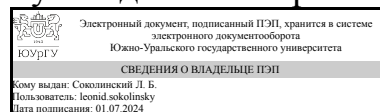


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Руководитель направления



Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.О.20 Интеллектуальный анализ данных
для направления 02.04.02 Фундаментальная информатика и информационные технологии

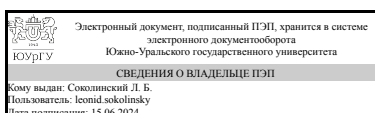
уровень Магистратура

форма обучения очная

кафедра-разработчик Системное программирование

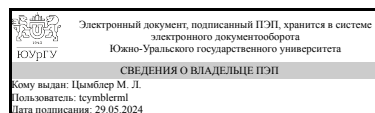
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 811

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



Л. Б. Соколинский

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



М. Л. Цымблер

1. Цели и задачи дисциплины

Целью дисциплины является ознакомление магистрантов с основными методами и алгоритмами интеллектуального анализа данных.

Краткое содержание дисциплины

Введение в дисциплину. Феномен Больших данных. Понятие интеллектуального анализа данных. Технологический цикл анализа данных. Основные 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, Eclat, FP-Growth. 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. Классификация. 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. Ансамблевая классификация: бэггинг, бустинг, случайный лес. Кластеризация. Задачи кластеризации данных и подходы к ее решению. Разделительная кластеризация: алгоритмы k-means, k-medoids и др. 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. Меры качества кластеризации: критерий Хопкинса, кросс-валидация, метод локтя, силуэтный коэффициент и др. Поиск аномалий. 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.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ОПК-2 Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	Знает: современные методы проектирования, разработки, отладки и тестирования приложений интеллектуального анализа данных Умеет: применять современные инструментальные средства для разработки приложений интеллектуального анализа данных Имеет практический опыт: применения современного программного инструментария для разработки приложений интеллектуального анализа данных

ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	Знает: методы подготовки данных и оценки эффективности моделей интеллектуального анализа данных Умеет: применять методы подготовки данных и оценки эффективности аналитических моделей для разработки приложений интеллектуального анализа данных Имеет практический опыт: применения программных средств для подготовки данных и оценки эффективности моделей интеллектуального анализа данных
ПК-1 Способен разрабатывать системы хранения и обработки больших данных, в том числе на основе методов искусственного интеллекта	Знает: определения, технологический цикл и основные методы решения базовых задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация, поиск аномалий) Умеет: выполнять проектирование приложений интеллектуального анализа данных Имеет практический опыт: разработки приложений интеллектуального анализа данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
ФД.02 Методы искусственного интеллекта, 1.О.03 Криптография и защита информации, 1.О.05 Объектно-ориентированные CASE-технологии, 1.О.17 Машинное обучение, 1.О.13 Программирование корпоративных информационных систем на языке Java, 1.О.09 Глубокие нейронные сети, ФД.01 Технологии интернета вещей, 1.О.12 Разработка систем искусственного интеллекта на языке Python	1.О.18 Языки разметки, 1.О.10 Технологии параллельного программирования, 1.О.07 Анализ информационных технологий, 1.О.21 Системы управления предприятием, 1.О.16 Квантовые вычисления

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

Дисциплина	Требования
ФД.02 Методы искусственного интеллекта	Знает: математические основы и технологии машинного обучения, современные интегрированные среды разработки ПО на языках высокого уровня и специализированные библиотеки искусственного интеллекта Умеет: применять современные методы машинного обучения на основе нейронных сетей, создавать и обучать глубокие и сверточные искусственные нейронные сети с применением специализированных библиотек Имеет практический опыт: анализа и оптимизации полученных решений на основе нейросетевого

	подхода, решения задач в области машинного обучения и компьютерного зрения
1.О.09 Глубокие нейронные сети	Знает: специализированные библиотеки для создания искусственных нейронных сетей, классы задач обработки больших данных на основе методов искусственных нейронных сетей, математическую модель нейрона, технологии создания искусственных нейронных сетей, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения Умеет: применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей, осуществлять формализацию задачи, построение математической модели, подготовку обучающего набора данных, подбор топологии и создание искусственной нейронной сети в соответствии с поставленной задачей Имеет практический опыт: создания и обучения искусственных нейронных сетей с применением специализированных библиотек, формулирования и решения задач в области машинного обучения с использованием нейросетевого подхода
ФД.01 Технологии интернета вещей	Знает: принципы организации киберфизических систем, существующие технологии в интернете вещей, отечественные и зарубежные достижения в области программно-аппаратных комплексов интернета вещей Умеет: анализировать существующие IoT-технологии и применять их в конкретных условиях, определять сервисы, функции и выбирать технологии их реализации при разработке киберфизических программно-аппаратных компонентов Имеет практический опыт: владения специальной терминологией, навыками программирования конечных устройств, навыками разработки моделей и алгоритмов для взаимодействия с программными и аппаратными компонентами, самостоятельного проектирования и реализации компонентов интернета вещей
1.О.12 Разработка систем искусственного интеллекта на языке Python	Знает: основные инструменты (программные библиотеки и язык программирования) для обработки и анализа данных, основные инструменты (программные библиотеки и язык программирования) для выполнения операций обработки и анализа данных, анализа готовых информационных наборов данных Умеет: подбирать наиболее подходящие инструменты для анализа имеющихся данных и выявления закономерностей, применять специализированные библиотеки языка Python для сбора, обработки и анализа данных, устанавливать программное обеспечение (среды разработок, программные библиотеки,

	соответствующий backend), просматривать версию и состав используемого программного обеспечения, задавать требуемый backend для решения поставленной задачи Имеет практический опыт: анализа готовых информационных наборов данных, применять специализированные библиотеки языка Python для сбора, обработки и анализа данных, установки и инсталляции программного обеспечения, используемого для решения задач в области сбора, обработки и анализа данных
1.О.13 Программирование корпоративных информационных систем на языке Java	Знает: основные понятия, виды и характеристики современного программного обеспечения технологии Java, основы объектно-ориентированного языка, основные понятия, виды и характеристики современного программного обеспечения технологии Java, подходы к тестированию программ на Java Умеет: использовать специализированные среды разработки Java, применять подходы объектно-ориентированного программирования при разработке программного обеспечения, проектировать и разрабатывать локальные приложения на языке Java, разрабатывать документацию с помощью Javadoc Имеет практический опыт: создания программных проектов в специализированных средах разработки Java, проектирования классов, ООП-архитектуры, создания программных проектов в специализированных средах разработки Java, разработки тестов для веб-сайта с помощью библиотеки Selenium
1.О.17 Машинное обучение	Знает: технологию создания моделей машинного обучения с помощью библиотек языка Python, методы оптимизации, регуляризации, нормализации и валидации моделей машинного обучения, математические основы, принципы создания, обучения и валидации моделей машинного обучения Умеет: создавать и обучать модели машинного обучения с помощью библиотек языка Python, математические основы, принципы создания, обучения и валидации моделей машинного обучения Имеет практический опыт: решения задач машинного обучения с помощью библиотек языка Python, анализа и оптимизации полученных решений на основе машинного обучения
1.О.03 Криптография и защита информации	Знает: основные подходы к математической формализации различных аспектов безопасности информационных систем и реализации средств защиты информации, основные требования информационной безопасности, основные алгоритмы шифрования данных, базовые понятия для математического обеспечения информационной безопасности Умеет: применять математические методы и алгоритмы

	защиты информации при решении профессиональных задач в области информационной безопасности, применять математические методы защиты информации, кодировать информацию с помощью основных алгоритмов шифрования Имеет практический опыт: самостоятельного формулирования задач и политик безопасности, построения систем защиты, использования основных алгоритмов шифрования для защиты данных и информационной безопасности
1.O.05 Объектно-ориентированные CASE-технологии	Знает: основные виды диаграмм UML, понятия, использующиеся в мета-языке UML и в конкретных видах диаграмм, основные особенности процесса проектирования программных систем, типы черт программных систем (поведенческие, структурные), классификацию моделей UML Умеет: строить модели проектируемого продукта с помощью различного типа диаграмм UML, выделять функциональные требования к разрабатываемой системе, определять поведенческие и структурные черты проектируемого программного обеспечения Имеет практический опыт: владения навыками анализа предметной области, спецификации поведенческих и структурных черт разрабатываемой информационной системы, оформления документации на этапе проектирования системы, владения навыками проектирования структуры и поведения программных систем

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 74,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	144	144
<i>Аудиторные занятия:</i>	64	64
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	69,5	69,5
Подготовка к контрольным опросам	69,5	69,5
Консультации и промежуточная аттестация	10,5	10,5
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен

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	8	0
3	Классификация	16	8	8	0
4	Кластеризация	16	8	8	0
5	Поиск аномалий	14	6	8	0

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

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	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
5	4	Разделительная кластеризация с помощью алгоритмов k-Means и k-Medoids.	2
6	4	Плотностная кластеризация с помощью алгоритма DBSCAN.	2
7	4	Иерархическая кластеризация с помощью различных мер схожести кластеров.	2
8	4	Вычисление качества кластеризации	2
9	5	Поиск точечных аномалий с помощью z-значимости, правила трех сигм, гистограмм.	4
10	5	Поиск коллективных аномалий метода вложенных циклов, метода решеток, кластеризации.	4

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Подготовка к контрольным опросам	Tan P.-N., Steinbach M., Karpatne A., Kumar V. Introduction to Data Mining. 2nd Edition. Pearson, 2019. 839 p.	2	69,5

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	2	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	экзамен

					минут.	
2	2	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск частых наборов"	3,5	10	экзамен

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;
 - 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	2	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск ассоциативных правил"	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; – 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	экзамен

						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	2	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.	экзамен
5	2	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 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	экзамен

					– 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	2	Текущий контроль	Практическое задание "Классификация. Деревья решений"	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; – 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	экзамен

						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	2	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 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.	экзамен

					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	2	Текущий контроль	Практическое задание "Классификация. Случайный лес"	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; – при исполнении кода на заданном	экзамен

					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	2	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 соответственно);	экзамен

					– 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	2	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.	экзамен
11	2	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	экзамен

					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; – 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	2	Текущий	Практическое	3,5	10	Выполнение задания оценивается от 0 экзамен

		контроль	задание "Кластеризация. Плотностная кластеризация"		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; – 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	2	Текущий контроль	Практическое задание "Кластеризация. Иерархическая 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 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:	экзамен

						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	2	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 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	экзамен

						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	2	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.	экзамен
16	2	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 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.	экзамен

					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	2	Текущий контроль	Практическое задание "Поиск аномалий. Поиск коллективных 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 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	экзамен

					– 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	2	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 один балл. Время на прохождение теста – не менее 15 минут.	экзамен
19	2	Промежуточная аттестация	Компьютерное тестирование	-	25	Промежуточная аттестация проводится во время экзамена в виде компьютерного теста. Тест состоит из 25 равноценных вопросов (под 5 вопросов на каждую из пяти тем курса), позволяющих оценить сформированность компетенций по курсу в целом, правильный ответ на один вопрос дает один балл. Время на прохождение теста – не менее 45 мин.	экзамен

6.2. Процедура проведения, критерии оценивания

1. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.

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
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Новикова О.А., Андрианова Е.Г. Анализ данных. М.: МИРЭА, 2020. 162 с. https://e.lanbook.com/book/167597
5	Дополнительная литература	Электронно-библиотечная система издательства Лань	Макшанов А.В., Журавлев А.Е. Технологии интеллектуального анализа данных. М.: Лань, 2019. 212 с. ISBN 978-5-8114-4493-9 https://e.lanbook.com/book/120063
6	Основная литература	Электронно-библиотечная система издательства Лань	Маккинни У. Python и анализ данных. М.: ДМК-Пресс, 2020. 540 с. https://e.lanbook.com/book/131721

Перечень используемого программного обеспечения:

Нет

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
Лекции		Проектор
Практические занятия и семинары		Персональный компьютер