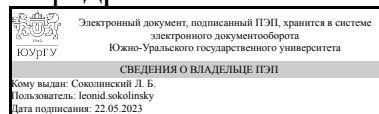


# ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:  
Заведующий выпускающей  
кафедрой



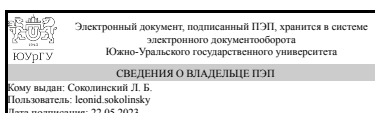
Л. Б. Соколинский

## РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П0.11.02 Технологии аналитической обработки информации  
для направления 09.03.04 Программная инженерия  
уровень Бакалавриат  
профиль подготовки Инженерия информационных и интеллектуальных систем  
форма обучения очная  
кафедра-разработчик Системное программирование

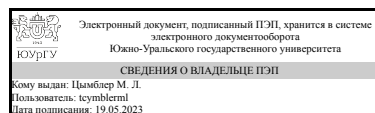
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.03.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 920

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 студентов с основными задачами оперативного и интеллектуального анализа данных и алгоритмами их решения.

### Краткое содержание дисциплины

1. Введение в дисциплину. Феномен Больших данных. Примеры предметных областей и задач, требующих аналитической обработки больших объемов данных. Понятия оперативного и интеллектуального анализа данных. Понятие хранилища данных. Технологический цикл аналитической обработки данных: интеграция источников данных, предварительная обработка данных, построение хранилища данных, интерпретация данных. Основные задачи интеллектуального анализа данных: поиск ассоциативных правил, классификация, кластеризация. 2. Хранилища данных и оперативный анализ данных. Понятие хранилища данных: предметная ориентированность, поддержка хронологии, интегрированность, неизменчивость. Отличия хранилищ данных и баз данных. Очистка данных: обработка отсутствующих и зашумленных данных. Интеграция, трансформация и редукция данных. Многомерная модель данных: измерения, меры, куб данных, OLAP-куб. Проектирование хранилищ данных: таблицы измерений, таблица фактов, схемы "звезда", "снежинка", "созвездие". Иерархии в измерениях. Технологический цикл построения хранилища данных, процессы ETL (Extract-Transform-Load). Понятие OLAP-куба. Обзор алгоритмов вычисления OLAP-куба: понятия полного куба, куба-айсберга, замкнутого куба и оболочки куба; методы многомерной агрегации и нисходящего вычисления подкубов. OLAP-операции: срез, вращение, агрегация, детализация. Расширения SQL ROLLUP BY и CUBE BY для вычисления OLAP-куба. Обзор современных систем создания и поддержки хранилищ данных. 3. Поиск ассоциативных правил. Понятия частого набора и ассоциативного правила. Алгоритм Apriori поиска частых наборов. Отбор ассоциативных правил на основе их поддержки и достоверности. 4. Классификация. Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Подходы к построению деревьев решений: индекс Джини, энтропия. Меры качества классификации: аккуратность, точность, полнота, F-мера. Методы ансамблевой классификации: бэггинг, бустинг, случайный лес. 5. Кластеризация. Разновидности подходов к кластеризации. Вычисление расстояния между кластеризуемыми объектами для различных видов их атрибутов (бинарные, интервальные, номинальные, порядковые и др.). Алгоритмы k-means и k-medoids. Плотностная кластеризация. Иерархическая кластеризация: дендрограммы, меры схожести кластеров (Single linkage, Complete linkage, Group average). Меры качества кластеризации: метод локтя, силуэтный коэффициент и др. Применение кластеризации для обнаружения аномалий в данных.

## 2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-5 (ПК-4 модели) Способен разрабатывать и применять методы машинного обучения для	Знает: методы и критерии оценки качества моделей машинного обучения

решения задач	Умеет: сопоставить задачам предметной области классы задач машинного обучения Имеет практический опыт: анализа требований и идентификации классов задач для реализации приложений машинного обучения
ПК-7 (ПК-8 модели) Способен разрабатывать системы анализа больших данных	Знает: общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных Умеет: настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных Имеет практический опыт: разработки программных компонент для извлечения и подготовки больших данных для аналитической обработки информации
ПК-11 (ПК-5 модели) Способен использовать инструментальные средства для решения задач машинного обучения	Знает: постановку базовых задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация) и базовые методы их решения Умеет: планировать и выполнять машинные эксперименты, оценивать точность и качество построенных моделей Имеет практический опыт: разработки приложений для аналитической обработки информации с помощью современных инструментальных средств

### 3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Теория вероятностей и математическая статистика, Основы машинного обучения, Базы данных	Основы разработки систем управления большими данными, Основы распределенной обработки данных, Современные языки программирования систем искусственного интеллекта

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

Дисциплина	Требования
Теория вероятностей и математическая статистика	Знает: основные математические положения, законы, основные формулы и методы решения задач теории вероятностей и математической статистики, статистические методы анализа данных Умеет: решать классические ( типовые) задачи теории вероятностей и математической статистики, применять математические методы для решения типовых профессиональных задач, ориентироваться в справочной литературе, применять знания из теории вероятностей и математической статистики для анализа данных

	<p>Имеет практический опыт: использования основных методов теории вероятностей и математической статистики для решения задач, связанных с профессиональной деятельностью, обработки данных средствами теории вероятностей и математической статистики</p>
Базы данных	<p>Знает: устройство интерфейсов между реляционными SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных, основы устройства систем баз данных, основы работы современных систем управления базами данных Умеет: использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа ) и операций с большими данными (например , матричные операции), установить и настраивать реляционные и нереляционные системы баз данных, создавать реляционные и нереляционные базы данных и запросы к ним Имеет практический опыт: написания запросов к реляционным и нереляционным большим базам данных, инсталляции систем баз данных, разработки реляционных и нереляционных баз данных</p>
Основы машинного обучения	<p>Знает: Знает классические методы и алгоритмы машинного обучения: предиктивные - обучение с учителем, дескриптивные - обучение без учителя , методы и инструментальные средства решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области, критерии выбора методов и инструментальных средств решения интеллектуальных задач, подходы к выбору методов и инструментальных средств систем искусственного интеллекта, процесс, стадии и методологии разработки решений на основе искусственного интеллекта, возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения Умеет: проводить сравнительный анализ и осуществлять выбор, настройку при необходимости разработку методов и алгоритмов для решения задач машинного обучения, осуществлять оценку критериев выбора методов и инструментальных средств решения задач с помощью систем искусственного интеллекта и выбор методов и инструментальных средств в зависимости от особенностей проблемной и предметной областей Имеет практический опыт: применения методов машинного обучения для решения задач, использования инструментальных средств решения задач искусственного интеллекта</p>

#### 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	
		6	
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	
Индивидуальное задание	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	Хранилища данных и OLAP	16	8	8	0
3	Поиск ассоциативных правил	14	6	8	0
4	Классификация	16	8	8	0
5	Кластеризация	16	8	8	0

### 5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Феномен Больших данных. Примеры предметных областей и задач, требующих аналитической обработки больших объемов данных. Понятия оперативного и интеллектуального анализа данных. Технологический цикл аналитической обработки данных: интеграция источников данных, предварительная обработка данных, построение хранилища данных, интерпретация данных. Основные задачи интеллектуального анализа данных: поиск ассоциативных правил, классификация, кластеризация.	2
2	2	Понятие хранилища данных: предметная ориентированность, поддержка хронологии, интегрированность, неизменчивость. Отличия хранилищ данных и баз данных. Очистка данных: обработка отсутствующих и зашумленных данных. Интеграция, трансформация и редукция данных. Многомерная модель данных: измерения, меры, куб данных, OLAP-куб. Проектирование хранилищ данных: таблицы измерений, таблица фактов, схемы "звезда", "снежинка", "созвездие". Иерархии в измерениях. Технологический цикл построения хранилища данных, процессы ETL (Extract-Transform-Load). Обзор современных систем создания и поддержки хранилищ данных.	4
3	2	Понятие OLAP-куба. Обзор алгоритмов вычисления OLAP-куба: понятия	4

		полного куба, куба-айсберга, замкнутого куба и оболочки куба; методы многомерной агрегации и нисходящего вычисления подкубов. OLAP-операции: срез, вращение, агрегация, детализация. Расширения SQL ROLLUP BY и CUBE BY для вычисления OLAP-куба.	
4	3	Понятия частого набора и ассоциативного правила. Антимонотонность поддержки. Алгоритм Apriori поиска частых наборов.	3
5	3	Отбор ассоциативных правил на основе их поддержки и достоверности. Компактное представление частых наборов: максимально частые и замкнутые наборы. Фрагментация и сэмплинг для поиска частых наборов	3
6	4	Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Подходы к построению деревьев решений: индекс Джини, энтропия. Меры качества классификации: аккуратность, точность, полнота, F-мера.	4
7	4	Ансамблевая классификация: бэггинг, бустинг, случайный лес.	4
8	5	Задачи кластеризации данных и подходы к ее решению. Разделительная кластеризация: алгоритмы k-means, k-medoids и др. Плотностная кластеризация: алгоритм DBSCAN, его применение для поиска аномалий в данных.	4
9	5	Иерархическая кластеризация: дендрограммы, агломеративный и дивизимный подход. Меры схожести кластеров: Single linkage, Complete linkage, Group average и др.	2
10	5	Меры качества кластеризации: критерий Хопкинса, кросс-валидация, метод локтя, силуэтный коэффициент и др.	2

## 5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	2	Разработка модельного хранилища данных: реализация очистки данных, реализация ETL приложения.	4
2	2	Разработка OLAP запросов к хранилищу данных.	4
3	3	Поиск частых наборов с помощью алгоритмов Apriori, ECLAT и FP-Growth.	4
4	3	Поиск ассоциативных правил с помощью мер support, confidence, lift.	4
5	4	Классификация с помощью дерева решений.	4
6	4	Ансамблевая классификация с помощью бэггинга. Ансамблевая классификация с помощью случайного леса. Ансамблевая классификация с помощью бустинга.	4
7	5	Разделительная кластеризация с помощью алгоритмов k-Means и k-Medoids.	2
8	5	Плотностная кластеризация с помощью алгоритма DBSCAN. Обнаружение аномалий с помощью кластеризации.	2
9	5	Иерархическая кластеризация с помощью различных мер схожести кластеров.	2
10	5	Вычисление качества кластеризации	2

## 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
Индивидуальное задание	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.	6	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	6	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.	экзамен
2	6	Текущий контроль	Практическое задание "Построение хранилища данных"	10	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 исполняется без синтаксических ошибок и предупреждений системы;	экзамен

					<p>– при выполнении кода на заданном наборе данных выдаются корректные результаты.</p> <p>2. Корректная визуализация полученных результатов.</p> <p>3. Понимание разработанного решения</p> <p>– студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</p> <p>4. Понимание полученных результатов</p> <p>– студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</p> <p>5. Готовность отчета</p> <p>– 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; пояснения, раскрывающие смысл полученных результатов.</p>		
3	6	Текущий контроль	Практическое задание "OLAP-запросы"	10	10	<p>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):</p> <p>1. Корректное решение поставленной задачи</p> <p>– представленный студентом программный код четко соответствует поставленной задаче;</p> <p>– код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно);</p> <p>– код выполняется без синтаксических ошибок и предупреждений системы;</p> <p>– при выполнении кода на заданном наборе данных выдаются корректные результаты.</p> <p>2. Корректная визуализация</p>	экзамен



					<p>полученных результатов.</p> <p>3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</p> <p>4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</p> <p>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; пояснения, раскрывающие смысл полученных результатов.</p>		
4	6	Текущий контроль	Контрольный опрос по теме "Хранилища данных и оперативный анализ данных"	4	10	<p>Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Хранилища данных и оперативный анализ данных".</p> <p>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.</p>	экзамен
5	6	Текущий контроль	Практическое задание "Поиск частых наборов"	12	10	<p>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):</p> <p>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</p>	экзамен

					<p>входных и выходных параметров соответственно);</p> <ul style="list-style-type: none"> <li>– код компилируется без синтаксических ошибок и предупреждений компилятора;</li> <li>– при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты.</li> </ul> <p>2. Корректная визуализация полученных результатов</p> <ul style="list-style-type: none"> <li>– представленный студентом код корректно строит диаграммы/графики, указанные в задании.</li> </ul> <p>3. Понимание разработанного решения</p> <ul style="list-style-type: none"> <li>– студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</li> </ul> <p>4. Понимание полученных результатов</p> <ul style="list-style-type: none"> <li>– студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</li> </ul> <p>5. Готовность отчета</p> <ul style="list-style-type: none"> <li>– 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.</li> </ul>		
6	6	Текущий контроль	Практическое задание "Поиск ассоциативных правил"	8	10	<p>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):</p> <p>1. Корректное решение поставленной задачи</p> <ul style="list-style-type: none"> <li>– представленный студентом программный код четко соответствует поставленной задаче;</li> </ul>	экзамен

					<p>– код документирован: файлы и подпрограммы исходного кода содержат спецификации (комментарии с описанием семантики кода в файле и описание семантики входных и выходных параметров соответственно);</p> <p>– код компилируется без синтаксических ошибок и предупреждений компилятора;</p> <p>– при исполнении кода на заданном наборе данных и с различными значениями параметров, указанными в задании, выдаются корректные результаты.</p> <p>2. Корректная визуализация полученных результатов</p> <p>– представленный студентом код корректно строит диаграммы/графики, указанные в задании.</p> <p>3. Понимание разработанного решения</p> <p>– студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</p> <p>4. Понимание полученных результатов</p> <p>– студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</p> <p>5. Готовность отчета</p> <p>– 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.</p>		
7	6	Текущий контроль	Контрольный опрос по теме "Поиск ассоциативных правил"	5	10	<p>Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Поиск ассоциативных правил".</p> <p>Прохождение компьютерного теста оценивается от 0 до 10 баллов. Тест состоит из 10 равнозначных вопросов, правильный ответ на один вопрос</p>	экзамен

					дает один балл. Время на прохождение теста – не менее 15 минут.	
8	6	Текущий контроль	Практическое задание "Классификация. Деревья решений"	5	10	экзамен

дает один балл. Время на прохождение теста – не менее 15 минут.

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.	
9	6	Текущий контроль	Практическое задание "Ансамблевая классификация"	15	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.		
10	6	Текущий контроль	Контрольный опрос по теме "Классификация"	5	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	6	Текущий контроль	Практическое задание "Кластеризация"	15	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

					<p>диаграммы/графики, указанные в задании.</p> <p>3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</p> <p>4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</p> <p>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.</p>		
12	6	Текущий контроль	Практическое задание "Качество кластеризации"	5	10	<p>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):</p> <p>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</p>	экзамен

					<p>результаты.</p> <p>2. Корректная визуализация полученных результатов – представленный студентом код корректно строит диаграммы/графики, указанные в задании.</p> <p>3. Понимание разработанного решения – студент в состоянии быстро и четко ответить на контрольные вопросы преподавателя, касающиеся разработанной программы.</p> <p>4. Понимание полученных результатов – студент в состоянии быстро и четко ответить на вопросы преподавателя, касающиеся содержательного смысла полученных результатов, включая их визуализацию.</p> <p>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.</p>		
13	6	Текущий контроль	Контрольный опрос по теме "Кластеризация"	5	10	<p>Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Кластеризация".</p> <p>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.</p>	экзамен
14	6	Промежуточная аттестация	Компьютерное тестирование	-	25	<p>Промежуточная аттестация проводится во время экзамена в виде компьютерного теста.</p> <p>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.</p>	экзамен



## 6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	<p>При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. № 25-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Отлично: Величина рейтинга обучающегося по дисциплине 85...100 %.</p> <p>Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %. Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 %. Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %.</p> <p>Если студент не согласен с оценкой, полученной по результатам текущего контроля, студент проходит мероприятие промежуточной аттестации в виде тестирования. Тестирование проводится в системе edu.susu.ru. Тест состоит из 25 равноценных вопросов, позволяющих оценить сформированность компетенций. На ответы отводится не менее 45 мин. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации.</p> <p>Фиксация результатов учебной деятельности по дисциплине проводится в день экзамена при личном присутствии студента.</p>	В соответствии с пп. 2.5, 2.6 Положения

## 6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
ПК-5	Знает: методы и критерии оценки качества моделей машинного обучения	+		++				++		+		+	+	+	
ПК-5	Умеет: сопоставить задачам предметной области классы задач машинного обучения	+		++				++		+		+	+	+	
ПК-5	Имеет практический опыт: анализа требований и идентификации классов задач для реализации приложений машинного обучения	+		++				++		+		+	+	+	
ПК-7	Знает: общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных	+	+	+	+	+	+	+	+	+	+	+	+	+	
ПК-7	Умеет: настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных	+	+	+	+	+	+	+	+	+	+	+	+	+	
ПК-7	Имеет практический опыт: разработки программных компонент для извлечения и подготовки больших данных для аналитической обработки информации	+	+	+	+	+	+	+	+	+	+	+	+	+	
ПК-11	Знает: постановку базовых задач интеллектуального анализа данных (поиск шаблонов, классификация, кластеризация) и базовые методы их решения	+			+	+	+	+	+	+	+		+	+	
ПК-11	Умеет: планировать и выполнять машинные	+			+	+	+	+	+	+	+		+	+	

	эксперименты, оценивать точность и качество построенных моделей																			
ПК-11	Имеет практический опыт: разработки приложений для аналитической обработки информации с помощью современных инструментальных средств	+			+	+	+	+	+	+	+	+	+	+	+	+			+	+

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. Методические указания к практическим занятиям
2. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.

*из них: учебно-методическое обеспечение самостоятельной работы студента:*

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. <a href="https://elibrary.ru/item.asp?id=43946965">https://elibrary.ru/item.asp?id=43946965</a>
2	Дополнительная литература	eLIBRARY.RU	Жаров А.Н., Минеичева И.Г. Анализ данных. Ярославль, 2020. 148 с. <a href="https://elibrary.ru/item.asp?id=43846458">https://elibrary.ru/item.asp?id=43846458</a>
3	Методические пособия для самостоятельной работы студента	Образовательная платформа Юрайт	Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с. <a href="https://urait.ru/bcode/432851">https://urait.ru/bcode/432851</a>

Перечень используемого программного обеспечения:

1. -Deductor Academic(бессрочно)

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