



Федеральное государственное бюджетное образовательное учреждение
высшего образования

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ

Кафедра Информационных систем и технологий

УТВЕРЖДАЮ
Начальник учебно-методического управления

«29» июня 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Основы анализа больших данных

направление подготовки/специальность 01.03.02 Прикладная математика и информатика

направленность (профиль)/специализация образовательной программы Прикладная математика и информатика

Форма обучения очная

Санкт-Петербург, 2023

1. Цели и задачи освоения дисциплины (модуля)

Целями освоения дисциплины «Основы анализа больших данных» являются: формирование у обучающихся системных фундаментальных знаний в области бизнес-аналитики, приобретение практических навыков использования методов аналитической обработки информации, применение на практике полученных знаний и умений в соответствии с международными требованиями к избранному виду деятельности.

Задачами освоения дисциплины являются:

- изучение технологий оперативного и интеллектуального анализа больших данных;
- ознакомление с базовыми понятиями и инструментарием информационно-аналитических систем;
- освоение методов анализа больших данных;
- освоение инструментария информационно-аналитических систем для решения задач анализа больших данных.

2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов освоения ОПОП
ПК-4 Способен осуществлять проектную деятельность, связанную с применением и разработкой математических моделей, алгоритмов, методов, программного обеспечения, в соответствии с поставленной задачей	ПК-4.2 Осуществляет выбор математических методов для разработки алгоритма решения поставленной задачи	знает – основные понятия определения Big Data; – методы извлечения знаний из данных. умеет – анализировать современные потоки данных; владеет – терминологией Data Mining;
ПК-4 Способен осуществлять проектную деятельность, связанную с применением и разработкой математических моделей, алгоритмов, методов, программного обеспечения, в соответствии с поставленной задачей	ПК-4.3 Осуществляет программную реализацию в соответствии с алгоритмом	знает – основные способы хранения и анализа данных; умеет – работать с программными средствами для хранения и анализа данных; владеет – навыками извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети).

3. Указание места дисциплины (модуля) в структуре образовательной программы

Данная дисциплина (модуль) включена в Блок «Дисциплины, модули» Б1.В.06 основной профессиональной образовательной программы 01.03.02 Прикладная математика и информатика и относится к части, формируемой участниками образовательных отношений учебного плана.

№ п/п	Предшествующие дисциплины	Код и наименование индикатора достижения компетенции
1	Теория вероятностей и математическая статистика	ОПК-2.1
2	Информационные технологии	УК-1.1, УК-1.2, УК-1.6
3	Программирование для ЭВМ	ОПК-5.1, ОПК-5.2

Для изучения дисциплины обучающиеся должны:

знать:

- особенности методов обработки информации, принципы работы с вычислительной техникой для решения задач обработки данных, основные понятия теории вероятностей и мат. статистики

уметь:

- пользоваться программным обеспечением для решения задач обработки данных, анализировать распределения случайных величин, исследовать законы распределений СВ.

владеть:

- навыками анализа данных в профессиональной сфере с использованием вычислительной техники и знаний ТВ и МС.

№ п/п	Последующие дисциплины	Код и наименование индикатора достижения компетенции
1	Выполнение и защита выпускной квалификационной работы	УК-1.1, УК-1.2, УК-1.3, УК-1.4, УК-1.5, УК-1.6, УК-2.1, УК-2.2, УК-2.3, УК-2.4, УК-3.1, УК-3.2, УК-3.3, УК-3.4, УК-3.5, УК-3.6, УК-4.1, УК-4.2, УК-4.3, УК-4.4, УК-5.1, УК-5.2, УК-5.3, УК-6.1, УК-6.2, УК-6.3, УК-6.4, УК-6.5, УК-7.1, УК-7.2, УК-7.3, УК-7.4, УК-8.1, УК-8.2, УК-8.3, УК-8.4, УК-9.1, УК-9.2, УК-9.3, УК-9.4, УК-9.5, УК-10.1, УК-10.2, УК-10.3, ОПК-1.1, ОПК-1.2, ОПК-1.3, ОПК-2.1, ОПК-2.2, ОПК-2.3, ОПК-3.1, ОПК-3.2, ОПК-4.1, ОПК-4.2, ОПК-4.3, ОПК-5.1, ОПК-5.2, ОПК-5.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-1.4, ПК-2.1, ПК-2.2, ПК-2.3, ПК-3.1, ПК-3.2, ПК-3.3, ПК-4.1, ПК-4.2, ПК-4.3, ПК-4.4, ПК(Ц)-1.1, ПК(Ц)-1.2, ПК(Ц)-1.3, ПК(Ц)-1.4, ПК(Ц)-1.5
2	Проектная практика	ПК-1.1, ПК-1.2, ПК-2.2, ПК-3.2, ПК-4.3, ПК(Ц)-1.4

4. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Вид учебной работы	Всего часов	Из них часы на практическую подготовку	Семестр
			7
Контактная работа	32		32
Практические занятия (Пр)	32	0	32
Иная контактная работа, в том числе:			
консультации по курсовой работе (проекту), контрольным работам (РГР)			
контактная работа на аттестацию (сдача зачета, зачета с оценкой; защита курсовой работы (проекта); сдача контрольных работ (РГР))			
контактная работа на аттестацию в сессию (консультация перед экзаменом и сдача			
Часы на контроль	4		4
Самостоятельная работа (СР)	36		36
Общая трудоемкость дисциплины (модуля)			
часы:	72		72
зачетные единицы:	2		2

5. Содержание дисциплины (модуля), структурированное по разделам (темам) с указанием отведенного на них количества академических часов и видов учебных занятий

5.1. Тематический план дисциплины (модуля)

№	Разделы дисциплины	Семестр	Контактная работа (по учебным занятиям), час.						СР	Всего, час.	Код индикатора достижения компетенции
			лекции		ПЗ		ЛР				
			всего	из них на практическую подготовку	всего	из них на практическую подготовку	всего	из них на практическую подготовку			
1.	1 раздел. Основы анализа больших данных (Big Data)										
1.1.	Введение в анализ больших данных	7			2			2	4	ПК-4.2, ПК-4.3	
1.2.	Инструментарий анализа больших данных	7			2			4	6	ПК-4.2, ПК-4.3	
1.3.	Предобработка данных	7			4			4	8	ПК-4.2, ПК-4.3	
2.	2 раздел. Методы интеллектуального анализа данных (Data mining)										
2.1.	Квантование данных. Кросс-таблица.	7			2			2	4	ПК-4.2, ПК-4.3	
2.2.	Классификация данных на основе деревьев решений	7			2			4	6	ПК-4.2, ПК-4.3	
2.3.	Кластеризация с использованием алгоритма k-means	7			4			4	8	ПК-4.2, ПК-4.3	
2.4.	Кластеризация на основе самоорганизующихся карт Кохонена	7			4			4	8	ПК-4.2, ПК-4.3	
3.	3 раздел. Модели и методы прогнозирования										
3.1.	Прогнозирование с использованием моделей линейной регрессии	7			4			4	8	ПК-4.2, ПК-4.3	
3.2.	Прогнозирование на основе пользовательской модели	7			4			4	8	ПК-4.2, ПК-4.3	
3.3.	Прогнозирование с помощью нейронной сети	7			4			4	8	ПК-4.2, ПК-4.3	
4.	4 раздел. Контроль										
4.1.	Зачет	7							4	ПК-4.2, ПК-4.3	

5.1. Практические занятия

№ разд	Наименование раздела и темы практических занятий	Наименование и содержание практических занятий
1	Введение в анализ больших данных	Введение в анализ больших данных Понятие больших данных. Методы анализа. Инструментарий решения задач в сфере BigData. Направления развития BigData

2	Инструментарий анализа больших данных	Аналитическая платформа Deductor Аналитическая платформа Deductor: возможности, начало работы, понятие сценария и узла обработки Базовые операции над узлами сценария. Мастер визуализации. Предобработка данных: парциальная обработка, выявление дубликатов и противоречий, спектральная обработка
3	Предобработка данных	Предобработка данных Предобработка данных: парциальная обработка, выявление дубликатов и противоречий, спектральная обработка, корреляционный анализ
4	Квантование данных. Кросс-таблица.	Квантование данных. Кросс-таблица. Квантование данных. Кросс-таблица. Классификация данных на основе деревьев решений. Кластеризация с использованием алгоритма k-means. Кластеризация на основе самоорганизующихся карт Кохонена
5	Классификация данных на основе деревьев решений	Классификация данных на основе деревьев решений Квантование данных. Кросс-таблица. Классификация данных на основе деревьев решений. Кластеризация с использованием алгоритма k-means. Кластеризация на основе самоорганизующихся карт Кохонена
6	Кластеризация с использованием алгоритма k-means	Кластеризация с использованием алгоритма k-means Квантование данных. Кросс-таблица. Классификация данных на основе деревьев решений. Кластеризация с использованием алгоритма k-means. Кластеризация на основе самоорганизующихся карт Кохонена
7	Кластеризация на основе самоорганизующихся карт Кохонена	Кластеризация на основе самоорганизующихся карт Кохонена Квантование данных. Кросс-таблица. Классификация данных на основе деревьев решений. Кластеризация с использованием алгоритма k-means. Кластеризация на основе самоорганизующихся карт Кохонена
8	Прогнозирование с использованием моделей линейной регрессии	Прогнозирование с использованием моделей линейной регрессии Модели линейной регрессии, предпосылки Гаусса-Маркова, оценка параметров модели, оценка качества модели. Прогнозирование с использованием моделей линейной регрессии.
9	Прогнозирование на основе пользовательской модели	Прогнозирование на основе пользовательской модели Прогнозирование с использованием линейной регрессии. Прогнозирование на основе пользовательской модели. Прогнозирование с помощью нейронной сети
10	Прогнозирование с помощью нейронной сети	Прогнозирование с помощью нейронной сети Прогнозирование с использованием линейной регрессии. Прогнозирование на основе пользовательской модели. Прогнозирование с помощью нейронной сети

5.2. Самостоятельная работа обучающихся

№ разд	Наименование раздела дисциплины и темы	Содержание самостоятельной работы
1	Введение в анализ больших данных	Введение в анализ больших данных Изучение теоретического материала
2	Инструментарий анализа больших данных	Аналитическая платформа Deductor Изучение теоретического материала. Выполнение практических заданий
3	Предобработка	Предобработка данных

	данных	Изучение теоретического материала. Выполнение практических заданий
4	Квантование данных. Кросс-таблица.	Квантование данных. Кросс-таблица. Изучение теоретического материала. Выполнение практических заданий
5	Классификация данных на основе деревьев решений	Классификация данных на основе деревьев решений Изучение теоретического материала. Выполнение практических заданий
6	Кластеризация с использованием алгоритма k-means	Кластеризация с использованием алгоритма k-means Изучение теоретического материала. Выполнение практических заданий
7	Кластеризация на основе самоорганизующихся карт Кохонена	Кластеризация на основе самоорганизующихся карт Кохонена Изучение теоретического материала. Выполнение практических заданий
8	Прогнозирование с использованием моделей линейной регрессии	Прогнозирование с использованием линейной регрессии Изучение теоретического материала. Выполнение практических заданий
9	Прогнозирование на основе пользовательской модели	Прогнозирование на основе пользовательской модели Изучение теоретического материала. Выполнение практических заданий
10	Прогнозирование с помощью нейронной сети	Прогнозирование с помощью нейронной сети Изучение теоретического материала. Выполнение практических заданий

6. Методические материалы для самостоятельной работы обучающихся по дисциплине (модулю)

Программой дисциплины предусмотрено проведение практических занятий, предполагающих закрепление изученного материала и формирование у обучающихся необходимых знаний, умений и навыков. Кроме того, важнейшим этапом изучения дисциплины является самостоятельная работа обучающихся с использованием всех средств и возможностей современных образовательных технологий.

В объем самостоятельной работы по дисциплине включается следующее:

- изучение теоретических вопросов по всем темам дисциплины;
- подготовка к практическим занятиям;
- подготовка к текущему контролю успеваемости студентов;
- подготовка к выполнению контрольной работы;
- подготовка к зачету.

Залогом успешного освоения дисциплины является обязательное посещение лекционных и лабораторных занятий, так как пропуск одного (тем более, нескольких) занятий может осложнить освоение разделов курса. На лабораторных занятиях материал, изложенный на лекциях, закрепляется при выполнении заданий.

Приступая к изучению дисциплины, обучающемуся необходимо в первую очередь ознакомиться с содержанием РПД, а также методическими указаниями по организации самостоятельной работы и подготовки к практическим занятиям.

При подготовке к лекционным занятиям студенту необходимо:

- ознакомиться с соответствующей темой занятия;
- осмыслить круг изучаемых вопросов и логику их рассмотрения;
- изучить рекомендуемую рабочей программой литературу по данной теме.

При подготовке к лабораторным занятиям и в рамках самостоятельной работы по изучению дисциплины обучающимся необходимо:

- повторить законспектированный на лекционном занятии материал и дополнить его с учетом рекомендованной по данной теме литературы;
- при самостоятельном изучении теоретической темы сделать конспект, используя рекомендованные в РПД источники;
- выполнить лабораторные задания в рамках изучаемой темы;
- ответить на контрольные вопросы по теме, используя материалы ФОС, либо групповые индивидуальные задания, подготовленные преподавателем;
- подготовиться к проверочной работе, предусмотренной в контрольных точках;
- подготовиться к промежуточной аттестации.

Итогом изучения дисциплины является зачет. Зачет проводится по расписанию. Форма проведения занятия может быть устная, письменная и в электронном виде. Студенты, не прошедшие аттестацию, должны ликвидировать задолженность в установленном порядке.

7. Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине (модулю)

7.1. Перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины

№ п/п	Контролируемые разделы дисциплины (модуля)	Код и наименование индикатора контролируемой компетенции	Вид оценочного средства
1	Введение в анализ больших данных	ПК-4.2, ПК-4.3	устный опрос
2	Инструментарий анализа больших данных	ПК-4.2, ПК-4.3	устный опрос, практические задания
3	Предобработка данных	ПК-4.2, ПК-4.3	устный опрос, практические задания
4	Квантование данных. Кросс-таблица.	ПК-4.2, ПК-4.3	устный опрос, тесты

5	Классификация данных на основе деревьев решений	ПК-4.2, ПК-4.3	устный опрос, тесты
6	Кластеризация с использованием алгоритма k-means	ПК-4.2, ПК-4.3	устный опрос, тесты
7	Кластеризация на основе самоорганизующихся карт Кохонена	ПК-4.2, ПК-4.3	устный опрос, тесты
8	Прогнозирование с использованием моделей линейной регрессии	ПК-4.2, ПК-4.3	устный опрос, практические задания
9	Прогнозирование на основе пользовательской модели	ПК-4.2, ПК-4.3	устный опрос, практические задания
10	Прогнозирование с помощью нейронной сети	ПК-4.2, ПК-4.3	устный опрос, практические задания
11	Зачет	ПК-4.2, ПК-4.3	

7.2. Типовые контрольные задания или иные материалы текущего контроля успеваемости, необходимые для оценки знаний, умений и навыков и (или) опыта профессиональной деятельности, характеризующих этапы формирования компетенций в процессе освоения дисциплины

(для проверки сформированности индикатора достижения компетенции ПК-4.2, 4.3)

Практические задания

1. Выполнить обзор инструментария для решения задач анализа больших данных

2. Подготовить доклад по теме,

примерные варианты тем:

Методы анализа больших данных решения задачи прогнозирования

Методы анализа больших данных решения задачи кластеризации данных

Методы анализа больших данных решения задачи классификации данных

Профессии в сфере анализа больших данных

Инструментарий инженера данных

Инструментарий Data Scientist

Инструментарий архитектора данных

и др.

3. Осуществить предварительную обработку данных (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц)

4. Построить уравнение линейной парной регрессии (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц). Оценить значимость коэффициентов, построить доверительные интервалы, построить прогноз, оценить прогноз, построить интервальный прогноз

5. Построить уравнение множественной регрессии (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц). Оценить значимость коэффициентов, построить доверительные интервалы, построить прогноз, оценить прогноз, построить интервальный прогноз

6. Построить дерево решений с использованием инструмента анализа (Python|R|иные инструменты АБД).

7. Решить задачу классификации объектов методом k-средних с использованием инструмента анализа (Python|R|иные инструменты АБД).

8. Решить задачу прогнозирования с использованием нейросетей с использованием инструмента анализа (Python|R|иные инструменты АБД).

9. Решить задачу прогнозирования с использованием деревьев решений с использованием инструмента анализа (Python|R|иные инструменты АБД).

Пример практического задания

Создайте BI-проект для анализа цен и тарифов транспортировки

Исходные данные для анализа хранятся в локальных таблицах:

1. Таблица «Справочник складов» (sklad_list.dbf) с полями: Код склада (CODE_SK), Номер склада (NUM_SK).

2. Таблица «Справочник клиентов» (client.dbf) с полями: Код клиента (CODE_CL),

Наименование клиента (NAME_CL), Тип договорных отношений (SALECOND).

3. Таблица «Поставки на склад» (sup_sklad.dbf) с полями: Дата поставки (DATA), Код документа (DOC_CODE), Код поставщика (SUP_CODE), Код склада (SKL_CODE), Код товара (W_CODE), Код транспорта (TRANS_CODE), Объем поставки (W_AMOUNT), Стоимость перевозки (TRANS_COST), Сумма поставки (SUM), Тариф на транспортировку (TRANS_TARI).

4. Таблица «Прайс-лист поставщиков» (price_list_sup.dbf) с полями: Единица измерения товара (DIM), Код поставщика (SUP), Код товара (CODE_W), Код транспорта (TRANS_CODE), Наименование товара (W_NAME), Тариф на транспортировку (TRANS_TARI), Тип товара (W_TYPE), Цена товара (PRICE).

На основании исходных данных необходимо спроектировать BI-проект, содержащий: OLAP-отчет, строящийся по запрашиваемым у пользователя типам товаров; Сценарий генерации микрокуба, содержащего данные о поставке товаров на

склады в первой половине ноября 2012 года. Создать сценарий массовой генерации микрокубов, каждый из которых содержит данные о поставках товаров одним видом транспорта. С помощью сгенерированных микрокубов необходимо решить следующие практические задачи:

1) Определить, какую долю транспортные расходы составляют от общей стоимости товара, поступившего на склады в первой половине ноября.

2) Определить динамику поступлений на склады оптовых партий обуви в первой половине ноября 2012 года.

3) Проанализировать, как распределилась доля товаров, поступивших на склады железнодорожным транспортом по типу договорных отношений с поставщиками. На какой склад поступило больше товаров по предоплате.

4) Проанализировать динамику доставки на склады железнодорожным транспортом оптовых партий одежды. Файл BI-проекта, файлы сценария, шаблоны микрокубов и результирующие микрокубы для проверки предъявить преподавателю.

7.3. Система оценивания результатов обучения по дисциплине (модулю) при проведении текущего контроля успеваемости

<p>Оценка «отлично» (зачтено)</p>	<p>знания:</p> <ul style="list-style-type: none"> - систематизированные, глубокие и полные знания по всем разделам дисциплины, а также по основным вопросам, выходящим за пределы учебной программы; - точное использование научной терминологии, систематически грамотное и логически правильное изложение ответа на вопросы; - полное и глубокое усвоение основной и дополнительной литературы, рекомендованной рабочей программой по дисциплине (модулю) <p>умения:</p> <ul style="list-style-type: none"> - умеет ориентироваться в теориях, концепциях и направлениях дисциплины и давать им критическую оценку, используя научные достижения других дисциплин <p>навыки:</p> <ul style="list-style-type: none"> - высокий уровень сформированности заявленных в рабочей программе компетенций; - владеет навыками самостоятельно и творчески решать сложные проблемы и нестандартные ситуации; - применяет теоретические знания для выбора методики выполнения заданий; - грамотно обосновывает ход решения задач; - безусловно владеет инструментарием учебной дисциплины, умение его эффективно использовать в постановке научных и практических задач; - творческая самостоятельная работа на практических/семинарских/лабораторных занятиях, активно участвует в групповых обсуждениях, высокий уровень культуры исполнения заданий
<p>Оценка «хорошо» (зачтено)</p>	<p>знания:</p> <ul style="list-style-type: none"> - достаточно полные и систематизированные знания по дисциплине; - усвоение основной и дополнительной литературы, рекомендованной рабочей программой по дисциплине (модулю) <p>умения:</p> <ul style="list-style-type: none"> - умеет ориентироваться в основных теориях, концепциях и направлениях дисциплины и давать им критическую оценку; - использует научную терминологию, лингвистически и логически правильно излагает ответы на вопросы, умеет делать обоснованные выводы; - владеет инструментарием по дисциплине, умение его использовать в постановке и решении научных и профессиональных задач <p>навыки:</p> <ul style="list-style-type: none"> - самостоятельная работа на практических занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий; - средний уровень сформированности заявленных в рабочей программе компетенций; - без затруднений выбирает стандартную методику выполнения заданий; - обосновывает ход решения задач без затруднений

<p>Оценка «удовлетворительно» (зачтено)</p>	<p>знания: - достаточный минимальный объем знаний по дисциплине; - усвоение основной литературы, рекомендованной рабочей программой; - использование научной терминологии, стилистическое и логическое изложение ответа на вопросы, умение делать выводы без существенных ошибок умения: - умеет ориентироваться в основных теориях, концепциях и направлениях по дисциплине и давать им оценку; - владеет инструментарием учебной дисциплины, умение его использовать в решении типовых задач; - умеет под руководством преподавателя решать стандартные задачи навыки: - работа под руководством преподавателя на практических занятиях, допустимый уровень культуры исполнения заданий; - достаточный минимальный уровень сформированности заявленных в рабочей программе компетенций; - испытывает затруднения при обосновании алгоритма выполнения решений</p>
<p>Оценка «неудовлетворительно» (не зачтено)</p>	<p>знания: - фрагментарные знания по дисциплине; - отказ от ответа (выполнения письменной работы); - знание отдельных источников, рекомендованных рабочей программой по дисциплине; умения: - не умеет использовать научную терминологию; - наличие грубых ошибок навыки: - низкий уровень культуры исполнения заданий; - низкий уровень сформированности заявленных в рабочей программе компетенций; - отсутствие навыков самостоятельной работы; - не может обосновать алгоритм выполнения заданий</p>

7.4. Теоретические вопросы и практические задания для проведения промежуточной аттестации обучающихся, необходимые для оценки знаний, умений и навыков и (или) опыта профессиональной деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

7.4.1. Теоретические вопросы для проведения промежуточной аттестации обучающихся

Примерный перечень вопросов:

1. Основные задачи, которые выполняют ИАС.
2. Роль и место анализа в принятии решений.
3. Проблемы анализа в свете использования информационных технологий.
4. Содержание аспекта сбора и хранения данных.
5. Содержание аспекта анализа данных и предоставления результатов анализа пользователям.

пользователям.

6. Классификация средств выполнения анализа с помощью ИТ.
7. Состав информационных технологий и информационных систем на предприятии
8. Понятие и структура информационного пространства.
9. Элементы структуры информационного пространства.
10. Понятия показателя и реквизитов.
11. Пространственная интерпретация понятия показатель.
12. Содержание экономических показателей.
13. Виды систем экономических показателей.
14. Рекомендации по структуризации информационного пространства предприятия при создании ИАС.

15. Содержание экономического анализа.
16. Сущность системы оценок.
17. Принципы гибкой архитектуры данных и открытых систем, которыми руководствуются при создании ИАС.
18. Информационный обмен, связанный с аналитической работой.
19. Понятие информационного хранилища.
20. Принципы построения информационных хранилищ.
21. Требования к качеству данных и способы его обеспечения при загрузке в информационное хранилище.
22. Концепции построения структур хранилищ данных.
23. Назначение, состав и выполняемые функции базы метаданных – репозитория ИХ
24. Принципы создания репозитория ИХ.
25. Элементы моделей данных ИХ (факт-таблица, таблицы измерений, консольные таблицы).
26. Схемы представления – модели многомерных данных.
27. Признаки OLAP-систем.
28. Типы многомерных OLAP-систем.
29. Классификация ИТ-анализа по режиму и темпу.
30. Задачи и содержание оперативного (OLAP) анализа.
31. Содержание понятия «знания», классификация видов знаний.
32. Интеллектуальный анализ данных (Data mining), цели и решаемые задачи.
33. Состав и содержание специфических задач интеллектуального анализа.
34. Классификация методов анализа.
35. Содержание методов анализа в экономической предметной области.
36. Состав программных инструментальных средств ИАС.
37. Средства сбора и доработки данных
38. Средства оперативного OLAP– анализа.
39. Средства интеллектуального анализа данных.
40. Понятие больших данных, 5V, 7V
41. Модели регрессии, оценка параметров, построение прогнозов
41. Деревья решений
42. Метод k-средних
43. Карты Кохонена
44. Прогнозирование на основе нейросетей

7.4.2. Практические задания для проведения промежуточной аттестации обучающихся (для проверки сформированности индикатора достижения компетенции ПК-4.2, ПК-4.3)

Практические задания

1. Осуществить предварительную обработку данных (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц)
2. Построить уравнение линейной парной регрессии (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц). Оценить значимость коэффициентов, построить доверительные интервалы, построить прогноз, оценить прогноз, построить интервальный прогноз
2. Построить уравнение множественной регрессии (по заданному варианту) с использованием инструмента анализа (Deductor|Python|R| электронных таблиц). Оценить значимость коэффициентов, построить доверительные интервалы, построить прогноз, оценить прогноз, построить интервальный прогноз
3. Построить дерево решений с использованием инструмента анализа (Python|R|иные инструменты АБД).
4. Решить задачу классификации объектов методом k-средних с использованием инструмента анализа (Python|R|иные инструменты АБД).
5. Решить задачу прогнозирования с использованием нейросетей с использованием инструмента анализа (Python|R|иные инструменты АБД).
6. Решить задачу прогнозирования с использованием деревьев решений с использованием инструмента анализа (Python|R|иные инструменты АБД).

7.4.3. Примерные темы курсовой работы (проекта) (при наличии)

Курсовая работа (проект) учебным планом не предусмотрена.

7.5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта профессиональной деятельности, характеризующие этапы формирования компетенций

Процедура проведения промежуточной аттестации и текущего контроля успеваемости регламентируется локальным нормативным актом, определяющим порядок организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся.

Процедура оценивания формирования компетенций при проведении текущего контроля приведена в п. 7.2.

Типовые контрольные задания или иные материалы текущего контроля приведены в п. 7.3.

Промежуточная аттестация по дисциплине проводится в форме зачета.

Зачет проводится в форме собеседования и практической части.

7.6. Критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии оценивания	Уровень освоения и оценка			
	Оценка «неудовлетворительно»	Оценка «удовлетворительно»	Оценка «хорошо»	Оценка «отлично»
	«не зачтено»	«зачтено»		
	Уровень освоения компетенции «недостаточный». Компетенции не сформированы. Знания отсутствуют, умения и навыки не сформированы	Уровень освоения компетенции «пороговый». Компетенции сформированы. Сформированы базовые структуры знаний. Умения фрагментарны и носят репродуктивный характер. Демонстрируется низкий уровень самостоятельности практического навыка.	Уровень освоения компетенции «продвинутый». Компетенции сформированы. Знания обширные, системные. Умения носят репродуктивный характер, применяются к решению типовых заданий. Демонстрируется достаточный уровень самостоятельности устойчивого практического навыка.	Уровень освоения компетенции «высокий». Компетенции сформированы. Знания аргументированные, всесторонние. Умения успешно применяются к решению как типовых, так и нестандартных творческих заданий. Демонстрируется высокий уровень самостоятельности, высокая адаптивность практического навыка

знания	<p>Обучающийся демонстрирует:</p> <ul style="list-style-type: none"> -существенные пробелы в знаниях учебного материала; -допускаются принципиальные ошибки при ответе на основные вопросы билета, отсутствует знание и понимание основных понятий и категорий; -непонимание сущности дополнительных вопросов в рамках заданий билета. 	<p>Обучающийся демонстрирует:</p> <ul style="list-style-type: none"> -знания теоретического материала; -неполные ответы на основные вопросы, ошибки в ответе, недостаточное понимание сущности излагаемых вопросов; -неуверенные и неточные ответы на дополнительные вопросы. 	<p>Обучающийся демонстрирует:</p> <ul style="list-style-type: none"> -знание и понимание основных вопросов контролируемого объема программного материала; - знания теоретического материала -способность устанавливать и объяснять связь практики и теории, выявлять противоречия, проблемы и тенденции развития; -правильные и конкретные, без грубых ошибок, ответы на поставленные вопросы. 	<p>Обучающийся демонстрирует:</p> <ul style="list-style-type: none"> -глубокие, всесторонние и аргументированные знания программного материала; -полное понимание сущности и взаимосвязи рассматриваемых процессов и явлений, точное знание основных понятий, в рамках обсуждаемых заданий; -способность устанавливать и объяснять связь практики и теории, -логически последовательные, содержательные, конкретные и исчерпывающие ответы на все задания билета, а также дополнительные вопросы экзаменатора.
умения	<p>При выполнении практического задания билета обучающийся продемонстрировал недостаточный уровень умений. Практические задания не выполнены. Обучающийся не отвечает на вопросы билета при дополнительных наводящих вопросах преподавателя.</p>	<p>Обучающийся выполнил практическое задание билета с существенными неточностями. Допускаются ошибки в содержании ответа и решении практических заданий. При ответах на дополнительные вопросы было допущено много неточностей.</p>	<p>Обучающийся выполнил практическое задание билета с небольшими неточностями. Показал хорошие умения в рамках освоенного учебного материала. Предложенные практические задания решены с небольшими неточностями. Ответил на большинство дополнительных вопросов.</p>	<p>Обучающийся правильно выполнил практическое задание билета. Показал отличные умения в рамках освоенного учебного материала. Решает предложенные практические задания без ошибок. Ответил на все дополнительные вопросы.</p>

владение навыками	Не может выбрать методику выполнения заданий. Допускает грубые ошибки при выполнении заданий, нарушающие логику решения задач. Делает некорректные выводы. Не может обосновать алгоритм выполнения заданий.	Испытывает затруднения по выбору методики выполнения заданий. Допускает ошибки при выполнении заданий, нарушения логики решения задач. Испытывает затруднения с формулированием корректных выводов. Испытывает затруднения при обосновании алгоритма выполнения заданий.	Без затруднений выбирает стандартную методику выполнения заданий. Допускает ошибки при выполнении заданий, не нарушающие логику решения задач Делает корректные выводы по результатам решения задачи. Обосновывает ход решения задач без затруднений.	Применяет теоретические знания для выбора методики выполнения заданий. Не допускает ошибок при выполнении заданий. Самостоятельно анализирует результаты выполнения заданий. Грамотно обосновывает ход решения задач.
-------------------	--	---	--	--

Оценка по дисциплине зависит от уровня сформированности компетенций, закрепленных за дисциплиной, и представляет собой среднее арифметическое от выставленных оценок по отдельным результатам обучения (знания, умения, владение навыками).

Оценка «отлично»/«зачтено» выставляется, если среднее арифметическое находится в интервале от 4,5 до 5,0.

Оценка «хорошо»/«зачтено» выставляется, если среднее арифметическое находится в интервале от 3,5 до 4,4.

Оценка «удовлетворительно»/«зачтено» выставляется, если среднее арифметическое находится в интервале от 2,5 до 3,4.

Оценка «неудовлетворительно»/«не зачтено» выставляется, если среднее арифметическое находится в интервале от 0 до 2,4.

8. Учебно-методическое и материально-техническое обеспечение дисциплины (модуля)

8.1. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины (модуля)

№ п/п	Автор, название, место издания, издательство, год издания учебной и учебно-методической литературы	Количество экземпляров/электронный адрес ЭБС
<u>Основная литература</u>		
1	Пальмов С. В., Интеллектуальный анализ данных, Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	http://www.iprbookshop.ru/75376.html
<u>Дополнительная литература</u>		
1	Федин Ф. О., Федин Ф. Ф., Анализ данных. Часть 2. Инструменты Data Mining, 2012	http://www.iprbookshop.ru/26445.html

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечиваются печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Наименование ресурса сети «Интернет»	Электронный адрес ресурса
--------------------------------------	---------------------------

Сайт разработчика программного обеспечения Deductor	https://basegroup.ru/
Онлайн курс "Big Data и Data Science: начни погружение с нуля"	https://stepik.org/course/101687/promo?search=2944259641
Онлайн курс "Big Data и Data Science: перейди на новый уровень"	https://stepik.org/course/101689/promo?search=2944259640

8.3. Перечень современных профессиональных баз данных и информационных справочных систем

Наименование	Электронный адрес ресурса
Система дистанционного обучения СПбГАСУ Moodle	https://moodle.spbgasu.ru/
Электронно-библиотечная система издательства "IPRsmart"	http://www.iprbookshop.ru/
Научная электронная библиотека eLIBRARY.RU	Научная электронная библиотека eLIBRARY.RU
Информационно-правовая система Гарант	\\law.lan.spbgasu.ru\GarantClient
Информационно-правовая система Консультант	\\law.lan.spbgasu.ru\ConsultantPlus ADM

8.4. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного и свободно распространяемого программного обеспечения

Наименование	Способ распространения (лицензионное или свободно распространяемое)
Microsoft Windows 10 Pro	Договор № Д32009689201 от 18.12.2020г
Deductor Academic версия 5.3	Свободно распространяемое
Python версия 3.7.6386.10	Свободно распространяемое
R версия 3.4.4	Свободно распространяемое
LibreOffice	Свободно распространяемое
PyCharm Community	Свободно распространяемое

8.5. Материально-техническое обеспечение дисциплины

Сведения об оснащённости учебных аудиторий и помещений для самостоятельной работы

Наименование учебных аудиторий и помещений для самостоятельной работы	Оснащённость оборудованием и техническими средствами обучения
73. Учебные аудитории для проведения практических занятий, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Учебная аудитория для проведения практических занятий, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации – комплект мультимедийного оборудования (персональный компьютер, мультимедийный проектор, экран, аудиосистема), доска маркерная белая эмалевая, комплект учебной мебели, подключение к компьютерной сети СПбГАСУ, выход в Интернет.
73. Помещения для самостоятельной работы	Помещение для самостоятельной работы (читальный зал библиотеки, ауд. 217): ПК-23 шт., в т.ч. 1 шт.- ПК для лиц с ОВЗ (системный блок, монитор, клавиатура, мышь) с подключением к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду СПбГАСУ. ПО Microsoft Windows 10

73. Компьютерный класс	Рабочие места с ПК (стол компьютерный, системный блок, монитор, клавиатура, мышь), стол рабочий, подключение к компьютерной сети СПбГАСУ, выход в Internet.
------------------------	---

Для инвалидов и лиц с ОВЗ обеспечиваются специальные условия для получения образования в соответствии с требованиями нормативно-правовых документов.