



Федеральное государственное бюджетное образовательное учреждение
высшего образования

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ

Кафедра Информатики

УТВЕРЖДАЮ
Начальник учебно-методического управления

«31» октября 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Визуализация данных в бизнес-анализе

направление подготовки/специальность 38.04.05 Бизнес-информатика

направленность (профиль)/специализация образовательной программы Бизнес-анализ систем
управления строительной организацией

Форма обучения очная

1. Цели и задачи освоения дисциплины (модуля)

Целью освоения дисциплины "Визуализация данных в бизнес-анализе" является изучение основных принципов визуализации данных и развитие навыков работы с инструментами визуализации.

Задачи дисциплины:

- освоение обучающимися популярных инструментов и программного обеспечения для визуализации данных;
- научить анализировать данные с помощью визуальных представлений, выявлять закономерности, тренды и аномалии.

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
ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.1 Применяет современные методы и программный инструментарий сбора, обработки и анализа данных	знает <ul style="list-style-type: none">– современные инструменты для обработки и анализа данных;– принципы статистического анализа и основные методы визуализации;– тенденции в области бизнес-анализа и визуализации данных; умеет <ul style="list-style-type: none">– собирать и интегрировать данные из различных источников;– применять статистические методы для анализа данных и выявления значимых закономерностей;– использовать программные инструменты для анализа и визуализации данных, включая создание интерактивных панелей и отчетов; владеет <ul style="list-style-type: none">– навыками работы с популярными инструментами визуализации данных;– навыками программирования на языке Python, используемом для анализа и визуализации данных.
ОПК-3 Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.2 Применяет интеллектуальное оборудование и системы искусственного интеллекта в процессе принятия управленческих решений	знает <ul style="list-style-type: none">– основные концепции и методы искусственного интеллекта и машинного обучения;– разнообразные методы визуализации, используемые для представления и вывода результатов моделей искусственного интеллекта; умеет <ul style="list-style-type: none">– определять подходящие инструменты и технологии моделей искусственного интеллекта для решения прикладных бизнес-задач;– визуализировать результаты анализа, выполненного с помощью искусственного интеллекта, для эффективного представления данных;

		– интерпретировать выводы моделей искусственного интеллекта и формировать на их основе рекомендательные отчеты для принятия управленческих решений; владеет - навыками работы с основными инструментами и библиотеками для машинного обучения; - умением создавать визуализации для интерпретации результатов работы моделей искусственного интеллекта, используя инструменты визуализации данных.
--	--	--

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

Данная дисциплина (модуль) включена в Блок «Дисциплины, модули» Б1.О.12 основной профессиональной образовательной программы 38.04.05 Бизнес-информатика и относится к обязательной части учебного плана.

№ п/п	Предшествующие дисциплины	Код и наименование индикатора достижения компетенции
1	Эконометрика	ОПК-1.1
2	Основы системного анализа и теории принятия решений	УК-1.1, УК-1.2, УК-1.3, УК-1.4, УК-1.5, УК-1.6

Эконометрика

Знать

- приемы выбора инструментальных средств для сбора и обработки статистической информации;

- общую процедуру применения методов сбора, обработки и анализа статистической информации при проведении эконометрического исследования;

- основные критерии оценки обработки и анализа собранной статистической информации при эконометрическом моделировании.

Уметь

- использовать современные компьютерные и информационные технологии при построении эконометрических моделей;

- собирать, обрабатывать и анализировать статистическую информацию при проведении эконометрического исследования ;

- обобщать и обосновывать полученные результаты исследования оформляя их в виде аналитического отчета, статистических таблиц и графиков;

Владеть:

- навыками работы с компьютерными программными продуктами при реализации методов сбора, обработки и анализа статистической информации.

Основы системного анализа и теории принятия решений

Знать

- теоретические основы теории проверки статистических гипотез;

Уметь

- формулировать статистические гипотезы;

- осуществлять процедуру проверки статистических гипотез;

Владеть

- процедурами проверки статистических гипотез.

№ п/п	Последующие дисциплины	Код и наименование индикатора достижения компетенции
1	Управление эффективностью бизнеса в строительстве	ПК-1.1, ПК-1.2, ПК-1.3
2	Управление корпоративными финансами	ПК-1.1, ПК-1.2, ПК-1.3

3	Подготовка к процедуре защиты 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-2.5, УК-3.1, УК-3.2, УК-3.3, УК-3.4, 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, ОПК-1.1, ОПК-1.2, ОПК-2.1, ОПК-2.2, ОПК-2.3, ОПК-3.1, ОПК-3.2, ОПК-3.3, ОПК-4.1, ОПК-4.2, ОПК-5.1, ОПК-5.2, ПК(Ц)-1.1, ПК(Ц)-1.2, ПК(Ц)-1.3, ПК(Ц)-1.4, ПК(Ц)-1.5, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-3.1, ПК-3.2, ПК-3.3, ПК-3.4
4	Управление инновационными проектами в сфере информационно-коммуникационных технологий	ОПК-2.1, ОПК-2.2, ОПК-2.3

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
			2
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)	72		72
Общая трудоемкость дисциплины (модуля)			
часы:	108		108
зачетные единицы:	3		3

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. Интеллектуальный анализ данных										
1.1.	Задача ассоциации в интеллектуальном анализе данных	2			8				18	26	ОПК-3.1, ОПК-3.2
1.2.	Нейронные сети в интеллектуальном анализе данных.	2			16				36	52	ОПК-3.1, ОПК-3.2
1.3.	Задача кластеризации в интеллектуальном анализе данных	2			8				18	26	ОПК-3.1, ОПК-3.2
2.	2 раздел. Контроль										
2.1.	Зачет	2								4	ОПК-3.1, ОПК-3.2

5.1. Практические занятия

№ разд	Наименование раздела и темы практических занятий	Наименование и содержание практических занятий
1	Задача ассоциации в интеллектуальном анализе данных	Задача ассоциации. Часть 1 Ассоциативные правила. Значимость ассоциативных правил. Интерпретация ассоциативных правил. Решение задачи ассоциации в АП Loginom.
1	Задача ассоциации в интеллектуальном анализе данных	Задача ассоциации. Часть 2 Решение задачи ассоциации средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи классификации в АП Loginom Особенности применения нейронной сети в качестве классификаторов. Подготовка исходных данных. Кодирование выходных значений. Выбор размера сети. Выбор архитектуры сети. Алгоритм построения классификатора.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи классификации средствами Python Решение задачи классификации средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи прогнозирования в АП Loginom Нейросеть (регрессия): настройка входных столбцов, настройка нормализации, разбиение на множества (метод разбиения, метод валидации), настройка параметров нейросети, настройка автоматического подбора параметров нейросети.

2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи прогнозирования средствами Python Решение задачи прогнозирования средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.
3	Задача кластеризации в интеллектуальном анализе данных	Решение задачи кластеризации в АП Loginom Алгоритм кластеризации k-means Алгоритм кластеризации G-means. EM кластеризация. Выбор числа кластеров.
3	Задача кластеризации в интеллектуальном анализе данных	Решение задачи кластеризации средствами Python Решение задачи кластеризации средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.

5.2. Самостоятельная работа обучающихся

№ разд	Наименование раздела дисциплины и темы	Содержание самостоятельной работы
1	Задача ассоциации в интеллектуальном анализе данных	Задача ассоциации Изучение введения в анализ ассоциативных правил: ассоциативные правила (association rules), обобщенные ассоциативные правила (generalized association rules), численные ассоциативные правила (quantitative association rules).
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи классификации в АП Loginom Нейронные сети в интеллектуальном анализе данных. Формальная модель нейрона. Нейронные сети.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи классификации средствами Python Подготовка к практическим занятиям. Решение задачи классификации средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи прогнозирования в АП Loginom Обучение персептрона. Выбор порога ошибки. Выбор числа итераций. Обучающая выборка. Ошибки при обучении Нейронной сети. Алгоритмы обучения персептрона.
2	Нейронные сети в интеллектуальном анализе данных.	Решение задачи прогнозирования средствами Python Подготовка к практическим занятиям. Решение задачи прогнозирования средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.
3	Задача кластеризации в интеллектуальном анализе данных	Решение задачи кластеризации в АП Loginom Изучение алгоритмов кластеризации в интеллектуальном анализе данных.
3	Задача кластеризации в интеллектуальном анализе данных	Решение задачи кластеризации средствами Python Подготовка к практическим занятиям. Решение задачи кластеризации средствами языка программирования Python с использованием библиотеки Plotly для визуализации данных.

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

Программой дисциплины предусмотрено проведение практических занятий, на которых дается основной систематизированный материал, а также выполняются практические задания, предполагающие 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 обучающихся;
- подготовка к зачету.

Залогом успешного освоения дисциплины является обязательное посещение практических занятий, так как пропуск одного (тем более, нескольких) занятий может осложнить освоение разделов курса. На практических занятиях теоретический материал закрепляется при выполнении практических заданий.

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

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

- ознакомиться с соответствующей темой занятия;
- осмыслить круг изучаемых вопросов и логику их рассмотрения;
- изучить рекомендуемую рабочей программой литературу по данной теме;
- при самостоятельном изучении теоретической темы сделать конспект, используя

рекомендованные в РПД источники;

- выполнить практические задания в рамках изучаемой темы;
- ответить на контрольные вопросы по теме, используя материалы ФОС;
- подготовиться к проверочной работе, предусмотренной в контрольных точках;
- подготовиться к промежуточной аттестации.

Работы, выполняемые на практических занятиях, сдаются только лично на занятиях преподавателю, который ведет группу. Задания, выполняемые на компьютере, обучающиеся сначала показывают только в электронном виде в соответствующих программах. При необходимости, при преподавателе доделывают или исправляют ошибки. Если требуется распечатать выполненные работы и сдать их в бумажном виде, преподаватель говорит об этом на занятиях.

Итогом изучения дисциплины является зачет. Зачет проводится по расписанию. Форма проведения занятия может быть устная, письменная и в электронном виде. Обучающиеся, 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	Задача ассоциации в интеллектуальном анализе данных	ОПК-3.1, ОПК-3.2	Устный опрос
2	Нейронные сети в интеллектуальном анализе данных.	ОПК-3.1, ОПК-3.2	Устный опрос
3	Задача кластеризации в интеллектуальном анализе данных	ОПК-3.1, ОПК-3.2	Устный опрос
4	Зачет	ОПК-3.1, ОПК-3.2	

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

Тестовые и контрольные задания для проверки сформированности индикаторов достижений компетенций ОПК-3.1, ОПК-3.2 размещены в Приложении.

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

<p>Оценка «отлично» (зачтено)</p>	<p>знания: - 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> <p>умения: - умеет ориентироваться в теориях, концепциях и направлениях дисциплины и давать им критическую оценку, используя научные достижения других дисциплин</p> <p>навыки: - 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; - безупречно 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>
<p>Оценка «хорошо» (зачтено)</p>	<p>знания: - достаточно полные и систематизированные знания по дисциплине; - усвоение основной и дополнительной литературы, рекомендованной рабочей программой по дисциплине (модулю)</p> <p>умения: - 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> <p>навыки: - 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>

<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 заданий</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

Примерные вопросы

1. Как формулируется задача ассоциации? Поясните структуру ассоциативного правила?
2. Что такое поддержка, достоверность и лифт правила?
3. Что называется популярным товарным набором?
4. На какие группы разделяются ассоциативные правила при их практической интерпретации?

Какие группы правил являются практически полезными?

5. По каким параметрам определяется практическая ценность ассоциативных правил?
6. Какой алгоритм используется для генерации ассоциативных правил? Какие параметры необходимо указать при его настройке?
7. Приведите математическую модель формального нейрона.
8. Почему сумматор вычисляет взвешенную сумму входных сигналов?
9. Зачем нейрону нужен порог активации? Что он определяет?
10. Какая функция активации используется в персептроне? Укажите ее вид.
11. Почему в персептронах не используются пороговые функции активации?
12. Поясните отличия между алгоритмом обратного распространения ошибки и алгоритмом L-BFGS.
13. Почему в AP Loginot для обучения персептрона используется алгоритм L-BFGS?

14. Опишите ошибки обучения нейронной сети.
15. Что такое переобучение сети? По какой причине оно возникает?
16. Перечислите правила формирования обучающей выборки.
17. Как определить количество нейронов в скрытом слое персептрона?
18. Каким способом можно оптимизировать структуру сети?
19. Поясните отличия постановки задачи кластеризации от задачи классификации.
20. Какие метрики используются в алгоритмах кластеризации?
21. Перечислите метрические алгоритмы кластеризации, которые реализованы в API Loginom.
22. Почему метрические алгоритмы плохо работают с категориальными признаками, измеренными в шкале наименований?
23. Чем метрика L1 отличается от метрики L2? В каких алгоритмах реализована каждая из них?
24. Перечислите эмпирические правила, которыми следует руководствоваться при выборе числа кластеров.
25. Укажите основные преимущества и недостатки EM алгоритма. Для решения каких задач его следует применять?
26. Сравните алгоритмы кластеризации k-means и G-means. В каких случаях следует отдать предпочтение каждому из них?
27. Почему результаты кластеризации в любом случае следует содержательно интерпретировать?

7.4.2. Практические задания для проведения промежуточной аттестации обучающихся

Практические задания для проведения промежуточной аттестации размещены в Приложении.

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

Курсовые работы (проекты) учебным планом не предусмотрены.

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. 7.3. Типовые контрольные задания или иные материалы текущего контроля приведены в п. 7.2. Промежуточная аттестация по дисциплине проводится в форме зачета. Зачет проводится в форме практического контрольного задания или тестового задания и собеседования.

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»		

	<p>Уровень освоения компетенции «недостаточный». Компетенции не сформированы. Знания отсутствуют, умения и навыки не сформированы</p>	<p>Уровень освоения компетенции «пороговый». Компетенции сформированы. Сформированы базовые структуры знаний. Умения фрагментарны и носят репродуктивный характер. Демонстрируется низкий уровень самостоятельности практического навыка.</p>	<p>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.</p>	<p>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</p>
знания	<p>Обучающийся демонстрирует: -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>	<p>Обучающийся демонстрирует: -знания теоретического материала; -неполные ответы на основные вопросы, ошибки в ответе, недостаточное понимание сущности излагаемых вопросов; -неуверенные и неточные ответы на дополнительные вопросы.</p>	<p>Обучающийся демонстрирует: -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>	<p>Обучающийся демонстрирует: -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.</p>

<p>умения</p>	<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>
<p>владение навыками</p>	<p>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.</p>	<p>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.</p>	<p>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.</p>	<p>Применяет теоретические знания для выбора методики выполнения заданий. Не допускает ошибок при выполнении заданий. Самостоятельно анализирует результаты выполнения заданий. Грамотно обосновывает ход решения задач.</p>

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	Маккинни У., Python и анализ данных, Москва: ДМК Пресс, 2020	https://e.lanbook.com/book/131721
2	Шарден Б., Массарон Л., Боскетти А., Крупномасштабное машинное обучение вместе с Python, Москва: ДМК Пресс, 2018	https://e.lanbook.com/book/105836
<u>Дополнительная литература</u>		
1	Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения, Москва: ДМК Пресс, 2017	https://e.lanbook.com/img/cover/book/100905.jpg
2	Волкова В. Н., Горелова Г. В., Лыпарь Ю. И., Паклин Н. Б., Фирсов А. Н., Черненькая Л. В., Козлов В. Н., Моделирование систем и процессов, Москва: Юрайт, 2024	https://urait.ru/bcode/536071
3	Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., Сиротин В. П., Анализ данных, Москва: Юрайт, 2024	https://urait.ru/bcode/536007

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
Система дистанционного обучения СПбГАСУ Moodle	https://moodle.spbgasu.ru/

8.3. Перечень современных профессиональных баз данных и информационных справочных систем

Наименование	Электронный адрес ресурса
Образовательные интернет-ресурсы СПбГАСУ	https://www.spbgasu.ru/university/obrazovatelnye-internet-resursy/
Российская государственная библиотека	www.rsl.ru
Научная электронная библиотека eLIBRARY.RU	https://www.elibrary.ru/
Электронно-библиотечная система издательства "IPRsmart"	http://www.iprbookshop.ru/
Электронно-библиотечная система издательства "ЮРАЙТ"	https://www.biblio-online.ru/
Электронно-библиотечная система издательства "Лань"	https://e.lanbook.com/
Электронная библиотека Ирбис 64	http://ntb.spbgasu.ru/irbis64r_plus/

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)
Python версия 3.7.6386.10	Свободно распространяемое
PyCharm Community	Свободно распространяемое

Loginom Community	Соглашение № 277/24 от 28.02.2024 г. с ООО «Аналитические технологии» на 6 лет
Anaconda	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
72. 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.
72. Компьютерный класс	Рабочие места с ПК (стол компьютерный, системный блок, монитор, клавиатура, мышь), стол рабочий, подключение к компьютерной сети СПбГАСУ, выход в Internet.
72. 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.

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

Рабочая программа составлена на основе ФГОС ВО - магистратура по направлению подготовки 38.04.05 Бизнес-информатика (приказ Минобрнауки России от 12.08.2020 № 990).

Программу составил:

доцент Инф, к.ф.-м.н. Мовсесова Л.В.

доцент, к.ф.-м.н. Ромаданова М.М.

Программа обсуждена и рекомендована на заседании кафедры Информатики
30.08.2024, протокол № 1

Заведующий кафедрой к.ф.-м.н. Л.В. Мовсесова

Программа одобрена на заседании учебно-методической комиссии факультета
19.09.2024, протокол № 2.

Председатель УМК д.э.н., профессор Г.Ф. Токунова

Проверочная работа

(для проверки сформированности индикатора достижения компетенции: ОПК-3.1, ОПК-3.2)

Задача ассоциации в АП Loginom

Постановка задачи. На основе исходного набора данных для выявления совместно приобретаемых товаров в супермаркетах выполните анализ рыночной корзины. В его основе лежит поиск ассоциативных правил, позволяющий находить закономерности между связанными событиями. Алгоритм может выявлять типичные шаблоны покупок и совместно приобретаемых товаров. Для поиска ассоциативных правил используется алгоритм FPG. Данные для анализа потребительской корзины предоставила розничная сеть, занимающаяся продажей бытовой химии. Набор данных насчитывает 5000 чеков. Данные представлены в текстовом файле.

Задача классификации в АП Loginom

Постановка задачи. На основе исходного набора данных риэлтерского агентства выполните классификацию недвижимости, продаваемую агентством и разделите ее по стоимости на 3 класса: «дешевые квартиры», «квартиры среднего сегмента» и «дорогие квартиры».

Задача прогнозирования в АП Loginom

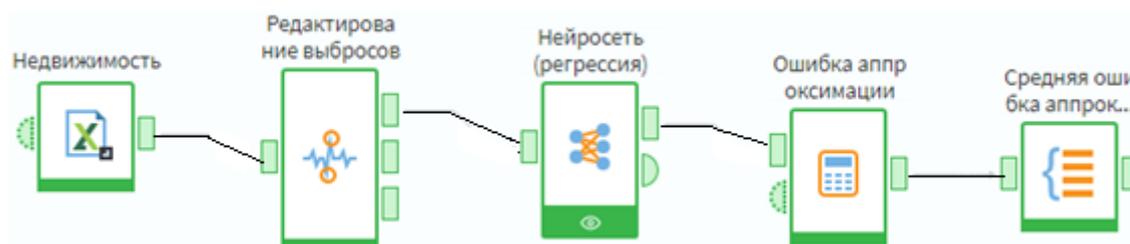
Постановка задачи. На рыночную стоимость квартир оказывают влияние многие факторы. Рынок недвижимости очень динамичный, поэтому требуется периодическая переоценка объектов собственности. Модели для оценки стоимости недвижимости могут существенно повысить эффективность работы риэлтерских организаций.

Структура исходных данных показана в табл.

Структура данных

Название поля	Тип	Вид
ID	Целый	Дискретный
Район	Текстовый	Дискретный
Количество комнат	Целый	Дискретный
Общая площадь, м2	Вещественный	Непрерывный
Жилая площадь, м2	Вещественный	Непрерывный
Площадь кухни, м2	Вещественный	Непрерывный
Состояние	Целый	Дискретный
Стоимость, т.руб.	Вещественный	Непрерывный

Сценарий прогнозирования стоимости недвижимости приведен на рисунке



1. В процессе ввода данных проконтролируйте тип и вид данных согласно таблице.

2. Подготовьте данные для анализа. Поля входного набора не должны содержать пропущенные значения. Для контроля пропущенных значений используйте визуализатор Статистика.
3. Настройка нейросети. Для решения задачи прогнозирования используется компонент Нейросеть (регрессия) .



Задача кластеризации в АП Loginom

Постановка задачи. На основе исходного набора данных с помощью подходящих узлов (Калькулятор, Преобразование полей, Фильтрация строк, Группировка) создать преобразованный набор, для которого выполнить кластеризацию. Провести интерпретацию полученных результатов. Дать пояснения по проведенным преобразованиям данных и отбору полей для проведения кластеризации. Построить для узла Кластеризация необходимые для описания полученных кластеров визуализаторы

Тестовые задания

(для проверки сформированности индикатора достижения компетенции: ОПК-3.1, ОПК-3.2)

Тема «Задача кластеризации в интеллектуальном анализе данных»

1. В кластер входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2,5,6,7. Расстояние от пятого объекта до кластера, если исходить из метода «ближнего соседа», равно:
 - а) 2;
 - б) 5;
 - в) 6;
 - г) 7.
2. В кластер входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2,5,6,7. Расстояние от пятого объекта до кластера, если исходить из принципа «дальнего соседа», равно:
 - а) 2;
 - б) 5;
 - в) 6;
 - г) 7.
3. В кластер входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2,5,6,7. Расстояние от пятого объекта до кластера, если исходить из принципа «средней связи», равно:
 - а) 2;
 - б) 5;
 - в) 6;

г) 7.

4. Процедура, при которой начальным является разбиение, состоящее из n одноэлементных классов, называется:

- а) дивизимной;
- б) корреляционной;
- в) агломеративной;
- г) линейной.

5. Кластерный анализ позволяет проводить:

- а) группировку объектов;
- б) группировку объектов и группировку признаков;
- в) группировку коэффициентов корреляции;
- г) группировку дисперсий;
- д) группировку признаков.

6. Цель кластерного анализа –

- а) образование групп сходных объектов;
- б) установление по результатам наблюдений адекватной аналитической зависимости между факторами;
- в) оценка степени тесноты связи между переменными;
- г) исследование влияния случайных величин на изменчивость факторов.

Тема «Классификация и регрессия»

1. Зависимость, при которой каждому значению величины X соответствует единственное значение величины Y и наоборот, называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

2. Зависимость, при которой каждому фиксированному значению независимой переменной X соответствует не одно, а множество значений переменной Y , называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

3. Зависимость, при которой функциональной зависимостью связаны фактор X и среднее значение результативного показателя Y , называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

4. Графическое изображение реальных статистических данных в виде точек в декартовой системе координат называется:

- а) диаграммой рассеяния;
- б) корреляционным полем;
- в) круговой диаграммой;
- г) верификацией модели.

5. Большинство программ, реализующих алгоритм иерархической классификации, предусматривает графическое представление результатов классификации в виде:

- а) дендрограммы;
- б) блок-схемы;
- в) дерева решений;
- г) ориентированного графа.

6. Цель корреляционного анализа в узком смысле –

- а) дать оценку степени тесноты связи между переменными;
- б) установить по результатам наблюдений адекватную аналитическую зависимость между факторами;
- в) исследовать влияние факторов на изменчивость средних значений наблюдаемых случайных величин;
- г) исследовать влияние случайных величин на изменчивость факторов.

7. Цель регрессионного анализа –

- а) установить по результатам наблюдений адекватную аналитическую зависимость между факторами;
- б) дать оценку степени тесноты связи между переменными;
- в) исследовать влияние факторов на изменчивость средних значений наблюдаемых случайных величин;
- г) исследовать влияние случайных величин на изменчивость факторов.

8. Главная цель метода канонических корреляций –

- а) поиск максимальных корреляций между группами факторных и результирующих признаков;
- б) установление по результатам наблюдений адекватной аналитической зависимости между факторами;
- в) оценка степени тесноты связи между переменными;
- г) исследование влияния случайных величин на изменчивость факторов.

Практические задания

При выполнении заданий необходимо произвести *предобработку данных*.

1. Загрузить набор данных (по своему варианту) как датафрейм библиотеки *pandas*.
2. Получить представления о наборе данных с помощью методов библиотеки *pandas* (*shape, head, tail, describe, info*). Посчитать количество пропущенных значений в признаках.
3. Произвести разведочный анализ данных – визуализировать данные, используя библиотеки *matplotlib, seaborn, plotly*.

Для этого построить (в зависимости от конкретного датасета и решаемой задачи):

- 1) парные диаграммы;

- 2) тепловую карту по матрице корреляции;
- 3) гистограммы;
- 4) диаграммы размаха («ящик с усами»);
- 5) графики распределения;
- 6) точечные 3D-графики распределения признаков.

4. Сделать выводы на основе проведенного анализа.

5. Произвести предобработку данных (в зависимости от решаемой задачи):

1) удалить из набора данных признаки, слабо коррелирующие между собой и с анализируемым признаком;

2) закодировать категориальные признаки в дискретные величины (метод *get_dummies* библиотеки *scikit-learn*);

3) объединить значения по периодам;

4) произвести нормализацию данных (метод *RobustScaler* библиотеки *scikit-learn*).

6. Разбить обработанный набор данных на обучающую и тестовую выборки (метод *train_test_split* библиотеки *scikit-learn*).

1. Задача бинарной классификации сбалансированного набора данных

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) метод опорных векторов (*Support vector machine*)

2) логистическая регрессия (*LogisticRegression*)

3) дерево решений (*Decision Tree*)

4) случайный лес (*RandomForest*)

– Отобразить корректность работы каждой модели на тестовой и обучающей выборках с помощью метрики *accuracy* библиотеки *scikit-learn*.

2. Задача многоклассовой классификации сбалансированного набора данных

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) метод опорных векторов (*Support vector machine*)

2) логистическая регрессия (*LogisticRegression*)

3) дерево решений (*Decision Tree*)

4) наивный Байес (*Naive Bayes*)

– Отобразить корректность работы каждой модели на тестовой и обучающей выборках с помощью метрики *accuracy* библиотеки *scikit-learn*.

3. Задача классификации несбалансированного набора данных

– Определить веса классов (*compute_class_weight*)

– Для учета несбалансированности набора данных использовать:

1) метод взвешенной классификации (*class_weight = 'balanced'*);

2) метод *oversampling* (*RandomOverSampler*);

3) метод *undersampling* (*RandomUnderSampler*).

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) Логистическая регрессия (*LogisticRegression*)

– Отобразить точность работы каждой модели на тестовой и обучающей выборках с помощью метрик *precision*, *recall*, *f1-score* и ROC-кривой библиотеки *scikit-learn*.

– Повторно отобразить метрики после проведения кросс-валидации.

4. Задача восстановления регрессии

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) линейная регрессия (*Linear Regression*);

2) регрессия LASSO (*Lasso Regression*);

3) гребневая регрессия (*Ridge Regression*);

4) полиномиальная регрессия (*Polynomial Regression*).

– Отобразить точность работы каждой модели на тестовой и обучающей выборках с помощью метрик *r squared* (коэффициент детерминации), *mean absolute error* (средняя абсолютная ошибка), *mean squared error* (средняя квадратичная ошибка) библиотеки *scikit-learn*.

– Повторно отобразить метрики после проведения кросс-валидации.

– Произвести поиск оптимальных гиперпараметров с помощью метода *GridSearchCV* по модели с наилучшими показателями.

5. Задача кластеризации

– Произвести снижение размерности набора данных

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) метод *k*-средних (*k-means clustering*)

– Произвести разведочный анализ данных по кластерам для оценки качества обучения модели.

6. Задача выявления аномалий

– Провести кластеризацию данных

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) метод *k*-средних (*k-means clustering*)

2) цепь Маркова (*Markov Chain*)

3) изолирующий лес (*Isolation forest*)

4) одноклассовый метод опорных векторов (*One class SVM*)

– Произвести разведочный анализ данных по набору данных для оценки качества обучения модели.

7. Задача прогнозирования временных рядов

– Провести кластеризацию данных

– Произвести обучение следующих моделей библиотеки *scikit-learn*:

1) *XGBoost Regressor*

– Отобразить точность работы модели на тестовой и обучающей выборках с помощью метрики *r squared* (коэффициент детерминации) библиотеки *scikit-learn*.