Приложение 4

к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.06 Анализ и визуализация данных
(индекс, наименование дисциплины в соответствии с учебным планом)

38.03.01 Экономика
(код, наименование направления подготовки)

Финансы и кредит
(наименование образовательной программы)

Очная, очно-заочная
(форма обучения)

Год набора – 2025

Чебоксары

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 00E49A235532F725E4778E40177F039EE7
Владелец: Ильина Любовь Николаевна
Действителен: с 08.05.2026 до 01.08.2027

Автор(ы)-составитель(и) РПД:

Скипина Людмила Николаевна, к.т.н., доцент кафедры экономики и менеджмента

Заведующий кафедрой:

Шигильчева Светлана Анатольевна, к.э.н., доцент, заведующий кафедрой экономики и менеджмента

Рабочая программа дисциплины Б1.В.06 Анализ и визуализация данных одобрена на заседании кафедры экономики и менеджмента Чебоксарского филиала РАНХиГС

протокол № 1 от «05» сентября 2024 г.

Актуализирована в соответствии с приказом ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» от 21 августа 2025 г. № 02-1592 «Об утверждении Положения о порядке разработки, утверждения и актуализации образовательных программ высшего образования – программам бакалавриата, специалитета, магистратуры»

протокол № 3 от «28» октября 2025 г.

Актуализирована в соответствии с приказом ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» от 29 мая 2026 г. № 02-1592 «Об актуализации рабочих программ дисциплин (модулей), практик, иных компонентов образовательных программ среднего профессионального образования и высшего образования» в части основной литературы в связи с прекращением с 01.07.2026 подписки на электронно-библиотечную систему «Юрайт»

протокол № 11 от «24» июня 2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы 4
2. Объем и место дисциплины (модуля) в структуре образовательной программы 6
3. Содержание и структура дисциплины (модуля) 7
4. Типы оценочных материалов, показатели и критерии оценивания 12
5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам 17
6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине 41
7. Методические материалы по освоению дисциплины (модуля) 59
8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет 61
9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы 63

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Дисциплина Б1.В.06 Анализ и визуализация данных обеспечивает формирование у обучающихся следующей профессиональной компетенции:

ОТФ/ТФ и реквизиты ПС *(при наличии)*	**Код компетенции**	**Наименование компетенции**	**Код индикатора достижения компетенций**	**Наименование индикатора достижения компетенций**	**Образовательный результат**
В Оценка влияния рисков на цели, ключевые показатели эффективности и решения, принимаемые в организациях B/03.5 Сбор данных, исследование и проверка данных для последующей оценки рисков организации ТД.1 Формирование списка источников исходных данных, обработка и подготовка исходных данных для последующей оценки влияния рисков на цели организации и бизнес-процессов (цели	ПКс ОС II - 6	Способен применять информационные технологии для решения прикладных финансово-экономических задач	ПКс ОС II - 6.1	Демонстрирует навыки использования информационных технологий для решения аналитических и управленческих задач	*ПКС ОС II - 6.1. З-1.* ***Знает*** источники данных для последующей оценки влияния рисков; *ПКС ОС II - 6.1. З-2.* ***Знает*** подходы к исследованию и методы исследования и проверки исходных данных для последующей оценки влияния рисков; *ПКС ОС II - 6.1. З-3.* ***Знает*** базовое программное обеспечение в области риск-ориентированного управления организацией, оценки влияния рисков на цели организации. *ПКС ОС II - 6.1. У-1.* ***Умеет*** формировать списки источников исходных данных; *ПКС ОС II - 6.1. У-2.* ***Умеет*** Обрабатывать и подготавливать исходные данные для последующей оценки влияния рисков на

конкретных управленческих решений, выполнение требований регуляторов и заинтересованных сторон) ТД.3 Формирование структурированных массивов с собранными данными ТД.4 Формирование списка ограничений, касающихся источников данных и качества исходных данных, для последующей оценки влияния рисков на цели организации и бизнес-процессов (цели конкретных управленческих решений, выполнение требований регуляторов и заинтересованных сторон), 08.018 Специалист по управлению рисками, утв. приказом Минтруда и социальной защиты РФ от					цели организации и бизнес-процессов (цели конкретных управленческих решений, выполнение требований регуляторов и заинтересованных сторон)

18 апреля 2025 г. N 264н					

2. Объем и место дисциплины (модуля) в структуре образовательной программы

Общий объем дисциплины:

4,00 з.е., 144 ак. часа

Контактная работа обучающихся с преподавателем по видам учебных занятий:

По очной форме обучения: 48 ак. часов на контактную работу с преподавателем, из них 16 ак. часов на лекции и 32 ак. часа на практические занятия. 92 ак. часа на самостоятельную работу обучающихся.

По очно-заочной форме обучения: 32 ак. часа на контактную работу с преподавателем, из них 8 ак. часов на лекции и 24 ак. часа на практические занятия. 108 ак. часов на самостоятельную работу обучающихся.

Б1.В.06 Анализ и визуализация данных реализуется на 4-м семестре 2 курса для студентов очной, очно-заочной ф/о после изучения дисциплин:

- Экономическая информатика
- Алгоритмизация и программирование.

3. Содержание и структура дисциплины (модуля)

3.1. Структура дисциплины (модуля)

Очная форма обучения

№ п/п	Наименование тем и (или) разделов		Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации
		ВСЕГО	Контактная работа обучающихся с преподавателем по видам учебных занятий									Самостоятельная работа		
			Период теоретического обучения						Период промежуточной аттестации (сессия)					
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Каттэк	Контроль	СРкр	СРэк	СР
			Л	ВЛ	ЛР	ПЗ								
Тема 1.	Подготовка данных к анализу	22	2	0	0	6	0	0	0	0		0	0	14	Опрос, практико-ориентированное задание
Тема 2.	Визуализация данных	24	4	0	0	4	0	0	0	0		0	0	16	Опрос, практико-ориентированное задание, контрольное задание
Тема 3.	Классификация и регрессия (Статистические	24	2	0	0	6	0	0	0	0		0	0	16	Опрос, практико-ориентированное задание

	методы)													
Тема 4.	Классификация и регрессия (Машинное обучение)	24	4	0	0	4	0	0	0	0	0	0	16	Тестирование, практико-ориентированное задание, контрольное задание
Тема 5.	Кластеризация	22	2	0	0	6	0	0	0	0	0	0	14	Опрос, практико-ориентированное задание
Тема 6.	Анализ и прогнозирование временных рядов	24	2	0	0	6	0	0	0	0	0	0	16	Опрос, практико-ориентированное задание
Промежуточная аттестация		4	0	0	0	0	0	0	0	4	0	0	0	Зачет с оценкой
Итого		144	16	0	0	32	0	0	4	4	0	0	92	

Очно-заочная форма обучения

№ п/п	Наименование тем и (или) разделов		Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации
		ВСЕГО	Контактная работа обучающихся с преподавателем по видам учебных занятий								Самостоятельная работа			
			Период теоретического обучения						Период промежуточной аттестации (сессия)					
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Каттэк	Контроль	СРкр	СРэк	СР
			Л	ВЛ	ЛР	ПЗ								
Тема 1.	Подготовка данных к анализу	24	2	0	0	4	0	0	0	0		0	0	18	Опрос, практико-ориентированное задание
Тема 2.	Визуализация данных	22	0	0	0	4	0	0	0	0		0	0	18	Опрос, практико-ориентированное задание, контрольное задание
Тема 3.	Классификация и регрессия (Статистические	24	2	0	0	4	0	0	0	0		0	0	18	Опрос, практико-ориентированное задание

	методы)													
Тема 4.	Классификация и регрессия (Машинное обучение)	22	0	0	0	4	0	0	0	0	0	0	18	Тестирование, практико-ориентированное задание, контрольное задание
Тема 5.	Кластеризация	24	2	0	0	4	0	0	0	0	0	0	18	Опрос, практико-ориентированное задание
Тема 6.	Анализ и прогнозирование временных рядов	24	2	0	0	4	0	0	0	0	0	0	18	Опрос, практико-ориентированное задание
Промежуточная аттестация		4	0	0	0	0	0	0	0	4	0	0	0	Зачет с оценкой
Итого		144	8	0	0	24	0	0	4	4	0	0	92	

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).

ИК – индивидуальные консультации.

КСР – контроль самостоятельной работы

КЭ – консультации перед экзаменом

Каттэк – контактная работа на аттестацию в период экзаменационных сессий

СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.

СРэк – самостоятельная работа на подготовку к экзамену.

СР – самостоятельная работа в семестре на подготовку к учебным занятиям.

3.2. Содержание дисциплины

Тема 1. Подготовка данных к анализу. ПКС ОС II - 6.1.

Очистка и предобработка данных.

Оценка качества данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений. Сокращение размерности. Сэмплинг.

Трансформация данных.

Группировка данных. Слияние данных. Квантование. Нормализация и кодирование.

Тема 2. Визуализация данных. ПКС ОС II - 6.1.

Введение в визуализацию. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа.

Визуализация данных средствами MS Power BI.

Семейство MS Power BI. Конструктор кубов и аналитических представлений. Совместная работа с представлениями – серверные сервисы Power BI. Решение задач объединения данных из разных источников, вычисляемые поля и Меры, элементы языка DAX, построение различных визуальных представлений по данным из учетных систем и CRM–систем.

Тема 3. Классификация и регрессия (Статистические методы). ПКС ОС II - 6.1.

Введение в классификацию и регрессию. Множественная линейная регрессия. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Нелинейные регрессионные модели. Основы логистической регрессии. Множественная логистическая регрессия. Оценка эффективности и сравнение моделей. ROC-анализ. Оценка ошибки модели. Издержки ошибочной классификации. LIFT– и PROFIT–кривые.

Тема 4. Классификация и регрессия (Машинное обучение). ПКС ОС II - 6.1.

Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений. Введение в нейронные сети. Принципы построения нейронных сетей. Процесс обучения нейронной сети. Обучение в условиях несбалансированности классов.

Тема 5. Кластеризация. ПКС ОС II - 6.1.

Введение в кластеризацию. Меры близости, используемые в алгоритмах кластеризации. Методы кластерного анализа. Иерархическая кластеризация. Алгоритм кластеризации k-means. Сети Кохонена. Карты Кохонена.

Тема 6. Анализ и прогнозирование временных рядов. ПКС ОС II - 6.1.

Введение в прогнозирование. Временной ряд и его компоненты. Модели прогнозирования. Исследование наборов данных и выбор метода прогнозирования. Ошибки прогноза. Оценка адекватности выбранного метода прогнозирования. Методы сглаживания и скользящие средние. Декомпозиция временного ряда. Регрессионный анализ временных рядов.

4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.В.06 Анализ и визуализация данных входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам (модулям) образовательной программы составляют фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа — это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа — это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	Ответ считается верным, если правильно указаны цифры или буквы
Задание закрытого типа с выбором нескольких правильных ответов из	Прочитайте текст, выберите правильные ответы	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.	Ответ считается верным, если правильно установлены все соответствия (позиции из одного

нескольких вариантов предложенных		2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).	столбца верно сопоставлены с позициями другого)
Задание закрытого типа на установление последовательности	Прочитайте текст и установите последовательность	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	Ответ считается верным, если правильно указана вся последовательность цифр
Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор	Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа

		ответа (например, 4 текст обоснования).	
Задание открытого типа с развернутым ответом	Прочитайте текст и запишите развернутый обоснованный ответ	1. Внимательно прочитать текст задания и понять суть вопроса. 2.Продумать логику и полноту ответа. 3.Записать ответ, используя четкие компактные формулировки. 4.В случае расчетной задачи, записать решение и ответ	Ответ считается верным: 1. Отсутствие фактических ошибок. 2. Раскрытие объема используемых понятий (полнота ответа). 3.Обоснованность ответа (наличие аргументов). 4. Логическая последовательность излагаемого материала.

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	Традиционная система	Бинарная система	ECTS	
			Для традиционной системы	Для бинарной системы
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100 баллов

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины Б1.В.06 Анализ и визуализация данных используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

опрос, практико-ориентированное задание, контрольное задание, тестирование.

5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных точек):

Тема 1. Подготовка данных к анализу. ПКс ОС II - 6.1.

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Для чего нужны очистка и предобработка данных.
2.	Оценка качества данных.
3.	Фильтрация данных.
4.	Обработка дубликатов и противоречий.
5.	Выявление аномальных значений.
6.	Восстановление пропущенных значений.
7.	Сокращение размерности. Сэмплинг.
8.	Трансформация данных.
9.	Группировка данных.
10.	Слияние данных.
11.	Квантование.
12.	Нормализация и кодирование.

Практико-ориентированное задание (компьютерный практикум):

Решение задачи мониторинга корректности консолидированных данных истории продаж сети предприятий розничной торговли

Для принятия управленческих решений о частоте проведения сезонных распродаж, которые должны увеличить объем выручки по итогам года, требуется анализ деятельности магазинов с возможностью прогнозирования выручки на следующий год. Для этого предоставляются данные о продажах 45 магазинов, расположенных в разных регионах - в каждом магазине есть несколько отделов. Компания также проводит несколько рекламных мероприятий по уценке в течение года, которые предшествуют знаменательным праздникам.

Табличные данные представлены в текстовых файлах в формате .csv:

1. stores.csv (магазины);
2. train.csv (продажи);
3. features.csv (признаки).

Магазины (stores.csv) — представлена анонимная информация о 45 магазинах (Store), с указанием типа (Type) и размера торговых площадей в

sf (Size) магазина.

Продажи (train.csv) — представлены реальные данные о продажах в сети Walmart, которые охватывают период с 2010 по 2012 года.

Признаки (features.csv) — представлены дополнительные данные, относящиеся к деятельности магазина, отдела и региона на указанные даты.

Выявление аномалий (также обнаружение выбросов) — это опознавание редких данных, событий или наблюдений, которые вызывают подозрения ввиду существенного отличия от большей части данных. Аномалии в данных могут возникать по разным причинам от неправильного ввода до ошибок оцифровки.

Анализ данных Stores

1. Преобразуйте категориальный признак Type в порядковый, так как в данном признаке есть скрытый порядок;

2. Преобразуйте признак Size в кв.м.;

3. Выявите аномалии (выбросы) в признаке Type и определите их в подходящую по значению к одному из типов магазинов, тем самым ослабив эффект выброса;

4. Сохраните полученные данные в файл: new _ stores . csv ;

5. Оцените качество очистки данных по каждому типу магазинов по признаку Size: соотнесите межквартильный размах до очистки данных к межквартильному размаху после очистки.

Внешнее обогащение предполагает привлечение дополнительной информации из внешних источников, что позволит повысить ценность и значимость данных с точки зрения их анализа. Под повышением значимости данных подразумевается, что на основе их анализа можно будет принимать управленческие решения принципиально нового уровня.

Анализ данных Train.

1. Выявите аномалии (выбросы) в признаке Weekly_Sales и преобразуйте их в подходящее значение, тем самым ослабив эффект выброса;

2. Сгруппируйте данные таким образом, чтобы получить новые данные следующего формата: сумма выручки по каждому магазину за определенный день;

3. Сформировать новую таблицу со следующими признаками: «Store», «Value», «Day», «Month», «Year», «IsHoliday»;

4. Сохраните полученные данные в файл: new_train.csv;

Зачастую данные, которые нам нужно использовать, имеют сложный вид; они не всегда бывают цельными, то есть разрозненный набор данных, который в процессе решения поставленной задачи требуется слить в один набор данных.

5. Соедините данные new_stores.csv и new_train.csv;

6. Сохраните полученные данные: value.csv;

7. Рассчитайте долю (%) магазинов сети Walmart каждого типа.

Внутреннее обогащение не предполагает привлечения какой-либо внешней информации.

Оно обычно связано с получением и включением в набор данных полезной информации, которая отсутствует в явном виде, но может быть тем или иным способом получена с помощью манипуляций с имеющимися данными.

Анализ данных Value.

Эффективность использования торговых площадей может оцениваться как по отделам внутри магазина, так и по магазину в целом.

1. Создайте новый признак Effic_space и рассчитайте по каждой позиции в данных эффективность использования одного квадратного метра торговой площади по каждому магазину (сколько выручки приходится на один квадратный метр торговой площади);

2. Сохраните полученные данные: new_value.csv;

3. Определите какой магазин эффективней всех использует торговую площадь, а какой неэффективно.

Шкалирование – это общепринятая задача предобработки данныхв машинном обучении. Многие алгоритмы машинного обучения – градиентные методы, которые сильно чувствительны к шкалированию. Они отлично работают на признаках, которые находятся на одной шкале, как правило, от 0 до 1 или от -1 до 1.

Анализ данных Features

Распространенный метод шкалирования данных является стандартизация, при котором значение признаков должны быть приближенно стандартно распределены.

1. Соединить данные: new_v alue.csv и features.csv;

2. Сделайте преобразования данных (аномалии, выбросы, категориальные признаки и т.п.);

3. Удалить признаки, которые дублируют друг друга или которые информативно включают в себя другие признаки.

После преобразования должны остаться следующие признаки: «Store», «Value», «IsHoliday», «Size», «Effic _ space», «Temperature», «Fuel _ Price», «MarkDown 2», «MarkDown 3», «MarkDown 4», «MarkDown 5», «CPI», «Unemployment» ;

4. Преобразуйте признаки, чтобы они имели среднее значение 0 и стандартное отклонение 1;

5. Определите, какой признак не требуется шкалировать;

6. Определите, какой признак имеет максимальное значение после шкалирования, а какой минимальный.

Тема 2. Визуализация данных. ПКс ОС II - 6.1

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Каковы цели и задачи визуализации данных в аналитических технологиях?
2.	Зачем нужен визуальный контроль данных после их загрузки в аналитическое приложение?
3.	Чем отличаются средства визуализации общего назначения от специализированных средств визуализации?
4.	Какие средства визуализации данных относят к визуализаторам общего назначения и почему?
5.	В чем заключается OLAP-анализ и каковы его цели?
6.	Какова структура OLAP-куба? Какие манипуляции с измерениями можно производить, чтобы сделать представление куба более информативным?
7.	Семейство MS Power BI. Конструктор кубов и аналитических представлений.
8.	Совместная работа с представлениями –серверные сервисы Power BI.
9.	Решение задач объединения данных из разных источников.
10.	Вычисляемые поля и Меры.
11.	Элементы языка DAX

Практико-ориентированное задание (компьютерный практикум):

Получение аналитической отчетности в MS Power BI

Легенда: аптечная сеть ООО «ФАРМА – Центр» имеет три торговых точки в разных городах России. Вы получили задание проанализировать работу фирмы за прошедший период и подготовить наглядный отчет.

Задача:

- Сформировать многомерный отчет, отражающий динамику сумм продаж по месяцам года в разрезе товарных групп.

- Сформировать многомерный отчет, отражающий динамику сумм продаж по месяцам года в разрезе аптек.

- Сформировать многомерный отчет и график продаж в регионах.

1. Откройте Power BI Desktop. Загрузите в Power BI данные для построения отчета. Тип данных — файлы Excel. Данные содержатся в файлах Продажи, Товары, Курс, География.

2. Объедините таблицы в единую модель данных.

3. Перейдите в представление Отчет. Используйте визуализатор «Карточка» для получения общей суммы продаж за период.

4. Постройте визуализатор «Гистограмма с группировкой», отметив поля Дата (таблица Курс) и Сумма (таблица Продажи). Получите представление по кварталам.

5. На панели Визуализации выберите Линейчатую диаграмму с группировкой

Отметьте поля Товарная группа (таблица Товары) и Сумма (таблица Продажи). Отсортируйте товарные группы по сумме выручки.

6. На панели Визуализации выберите Круговую диаграмму.

Отметьте поля Область (таблица География), Отдел и Сумма (таблица Продажи).

7. На панели Визуализации выберите диаграмму Карта. Отметьте поля Город (таблица География) и Сумма (таблица Продажи). В области форматирования удалите название диаграммы. Параметр «Пузырьковая» установите — 10%. Измените цвета данных и стили карт.

8. Добавьте в отчет новую страницу. Новую страницу переименуйте в Тренд. На странице Тренд постройте визуализатор График. Отметьте поля Дата (таблица Курс) и Сумма (таблица Продажи). Добавьте Линию тренда.

9. На панели Визуализации выберите Срез: поле Товарная группа (таблица Товары).

10. Добавьте срез по городам.

Контрольное задание:

1. Из нескольких открытых источников подобрать данные, которые позволят

– дать одного всестороннюю характеристику из регионов РФ;

− проанализировать динамику основных показателей социально-

экономического развития региона.

2. Загрузить данные в MS Power BI, построить модель данных, провести необходимые преобразования данных, вычислить меры.

3. Подобрать средства визуализации данных. Построить отчет, который позволит проанализировать демографическое, социальное, экономическое положение региона РФ и динамику основных показателей.

4. Подготовить презентацию с использованием MS Sway для иллюстрации полученных результатов.

Тема 3. Классификация и регрессия (статистические методы). ПКс ОС II - 6.1

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	В чем заключается задача классификации?
2.	Какие алгоритмы используются для решения задач классификации?
3.	По какому принципу строится таблица сопряженности и каковы цели ее использования?
4.	Как рассчитываются показатели чувствительности и специфичности модели, в чем состоит их смысл?
5.	Как строится ROC-кривая?
6.	Как можно определить оптимальный порог отсечения?
7.	Как строится LIFT-кривая?
8.	Какие достоинства и недостатки имеют модели логистической регрессии применительно к задаче кредитного скоринга?

Практико-ориентированное задание (компьютерный практикум):

Скоринговые модели для оценки кредитоспособности заемщиков

Постановка задачи.

В коммерческом банке имеется продукт «Нецелевой потребительский кредит»: кредиты предоставляются на любые цели с принятием решения в течение нескольких часов. За это время проверяются минимальные сведения о клиенте, в основном, такие, как отсутствие криминального прошлого и

кредитная история в других банках.

В банке накоплена статистическая информация о заемщиках и качестве обслуживания ими долга за несколько месяцев. Руководство банка, понимая, что отсутствие адекватных математических инструментов, позволяющих оптимизировать риски, не способствует расширению розничного бизнеса в области потребительского кредитования, поставило перед отделом розничных рисков задачу разработать скоринговые модели с различными стратегиями кредитования, которые позволили бы управлять рисками и минимизировать число «безнадежных» заемщиков.

Решение задачи.

Импортировать файл с кредитными историями Loans.txt.

В задаче представлено 2709 кредитов с известными исходами платежей. Набор данных уже разбит на два множества — обучающее (80%) и тестовое (20%) — таким образом, чтобы в каждом множестве доля «плохих» кредитов была примерно одинаковой.

В структуре данных:

Поле	Описание
Код	Служебный код заявки
Дата	Дата выдачи кредита
О/Д, %	Коэффициент Обязательства/Доход в процентах (Вещественный)
Возраст	Возраст заемщика на момент принятия решения о выдачи кредита
Проживание	Основание для проживания: собственник, муниципальное жилье, аренда
Срок проживания в регионе	Менее 1 года, от 1 года до 5 лет, свыше 5 лет
Семейное положение	Холост/не замужем, женат/замужем, разведен/вдовство, другое
Образование	Среднее, среднее специальное, высшее
Стаж работы на последнем месте	Менее 1 года, от 1 года до 3 лет, свыше 3 лет
Уровень должности	Сотрудник, руководитель среднего звена, руководитель высшего звена
Кредитная история	Информация из бюро кредитных историй. Если есть информация о просрочках по прошлым кредитам, то клиенту присваивается категория «отрицательная» (тип данных Строковый);
Просрочки свыше 60 дней	0 – отсутствовали, 1 – имели место (тип данных Целый);
Тестовое множество	Служебный признак, TRUE соответствует тестовому множеству (тип данных Логический).

1. С помощью обработчика «Калькулятор» на основе поля Просрочки более 60 дн. (COL12) получите новое поле Класс заемщика.

2. Постройте модель Логистической регрессии. Выходная переменная

— Класс заемщика.

3. Запишите уравнение логистической регрессии.

4. Используя визуализатор «Таблица сопряженности», определите Чувствительность и Специфичность полученной модели.

5. Оцените качество модели.

6. Чему равен оптимальный порог отсечения?

— Определить оптимальный порог отсечения, используя визуализатор ROC-кривая.

— Определить оптимальный порог отсечения, используя правило

$$P = \frac{1}{\left(1 + {C_{FN}}/{C_{FP}}\right)}$$

C_{FN} — издержки, связанные с ложноотрицательными наблюдениями;

C_{FP} — издержки, связанные с ложноположительными наблюдениями.

Отношение издержек ${C_{FN}}/{C_{FP}}$ задать самостоятельно.

7. Как изменяется кредитный рейтинг заемщика с каждым дополнительным иждивенцем?

8. Используя визуализатор «Что–если», определите к какому классу будут отнесены заемщики

— Мужчина, 43-х лет, состоит в браке, 2 иждивенца, доход — 50000 руб., опыт работы — 20 лет, срок проживания на одном месте —10 лет., недвижимость 0, месячный платеж 7500 руб.

— Женщина, 30-ти лет, не замужем, 1 иждивенец, доход — 40000 руб., опыт работы — 6 лет, срок проживания — 10 лет, недвижимость 0, месячный платеж 5000 руб.

Тема 4. Классификация и регрессия (Машинное обучение). ПКс ОС II - 6.1

Тестовые задания:

Тест 1.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать несколько правильных ответов.

Записать только буквы выбранных вариантов ответа.

Классификация — это ...

a. отнесение объектов к одной из заранее неизвестных групп
b. процесс формирования групп и отнесения объектов к одному из них.
c. отнесение объектов к одному из заранее известных классов

Задачи классификации решаются следующими алгоритмами:
a. алгоритмом Apriori
b. нейронные сети
c. линейной регрессии

Деревья решений относятся к группе (-ам) …
a. кибернетических методов
b. методов кросс-табуляции
c. статистических методов
d. логических методов

Внутренний узел дерева решений называют также …
a. узлом проверки
b. конечным узлом
c. вершиной
d. листом

Конечный узел дерева решений называют также ...
a. узлом проверки
b. узлом решения
c. листом

При помощи метода деревьев решений решаются задачи …
a. классификации и кластеризации
b. классификации и прогнозирования
c. кластеризации, классификации и прогнозирования
d. кластеризации и прогнозирования

Процесс отсечения ветвей или замена некоторых ветвей поддеревом …
a. происходит сверху вниз, т.е. является нисходящим
b. происходит снизу вверх, т.е. является восходящим
c. может быть, как нисходящим, так и восходящим

Алгоритм конструирования дерева решений …
a. не требует от пользователя выбора из набора входных атрибутов

(независимых переменных), наиболее значимых

b. требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых

c. на вход алгоритма можно подавать все существующие атрибуты, алгоритм сам выберет наиболее значимые среди них, и только они будут использованы для построения дерева

Если зависимая переменная принимает непрерывные значения, то дерево решений решает задачу:

a. классификации и численного прогнозирования

b. численного прогнозирования

c. классификации

Если зависимая переменная принимает дискретные значения, при помощи метода дерева решений решается задача:

a. классификации и численного прогнозирования

b. классификации

c. численного прогнозирования

Метод, который делает заключения относительно данной ситуации по результатам поиска аналогий, хранящихся в базе прецедентов относится к категории …

a. обучение с учителем"

b. самообучающейся системы

c. "обучение без учителя"

Явление переобучения характеризуется …

a. чрезмерно точным соответствием нейронной сети конкретному набору обучающих примеров, при котором сеть теряет способность к обобщению

b. возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры нейронной сети

c. возникновением, в случае слишком долгого обучения, слишком большого числа обучающих примеров или слишком сложной структуры нейронной сети

Многослойный персептрон – это сеть …

a. обратного распространения сигнала (с обратными связями)

b. в которой входной сигнал преобразуется в выходной, проходя последовательно через несколько слоев

c. прямого распространения сигнала (без обратных связей)

В многослойном персептроне …

a. должен быть хотя бы один скрытый слой

b. присутствие нескольких скрытых слоев оправдано лишь в случае использования нелинейных функций активации

c. может быть какое угодно количество скрытых слоев, они также могут вообще отсутствовать

Заполните пропуски в формулировке: "… – выходная связь нейрона, с которой сигнал (возбуждения или торможения) поступает на… следующих нейронов"

a. синапс, аксоны

b. аксон, точку ветвления

c. аксон, синапсы

Однонаправленные входные связи, соединенные с выходами других нейронов – это …

a. синапсы

b. слои сети

c. аксоны

Слой нейронной сети – это …

a. выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов

b. один или несколько нейронов, на входы которых подается один и тот же общий сигнал

c. однонаправленные входные связи, соединенные с выходами других нейронов

Ошибкой обучения нейронной сети называется …

a. переобучение нейронной сети

b. разность между желаемым и полученным на выходе сигналами

c. целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети

Целевая функция, требующая минимизации в процессе управляемого

обучения нейронной сети – это …

a. функция ошибок

b. функция переобучения

c. ошибка обучения

Практико-ориентированное задание (компьютерный практикум):

Прогнозирование ухода к другому оператору клиентов телекоммуникационной сети

Предлагается решить задачу из области прогнозирования ухода к другому оператору клиентов телекоммуникационной сети.

В двух файлах telcojan.txt (обучающая выборка) и telcofeb.txt (тестовая выборка) имеются статистические данные клиентах телекоммуникационной компании по состоянию на январь и февраль некоторого года.

Структура файлов одинакова и содержит уже отобранные аналитиками наиболее значимые характеристики, влияющие на уход клиента (отказ от пользования услугами): демографическая информация и используемый тариф.

N п/п	**Имя поля**	**Описание**
1	Срок клиента	Количество месяцев, прошедших с момента подключения
2	Возраст	Возраст клиента
3	Адрес	Срок проживания по последнему адресу
4	Стаж	Стаж работы общий
5	Семейное положение	Семейное положение клиента
6	Образование	Уровень образования
7	Пенсионер	–
8	Пол	–
9	Тариф	Текущий тариф клиента
10	Уход	Отказ от пользования услугами (соответствует значению 1)

1. Постройте модель «Дерево решений».

2. Используя визуализатор «Таблица сопряженности», определите

Чувствительность и Специфичность полученной модели.

3. Оцените качество модели. Сравните ее с моделью, полученной на предыдущем этапе.

4. С помощью визуализатора «Значимость атрибутов» охарактеризуйте, степень нелинейной зависимости между каждым фактором, включенным в модель, и независимой переменной.

5. Постройте модель «Нейронная сеть». Используйте алгоритм BACKPROP.

7. Используя визуализатор Таблица сопряженности, определите Чувствительность и Специфичность полученной модели.

8. Оцените качество модели. Сравните ее с моделью, полученной на предыдущем этапе.

9. Переобучите нейронную сеть

– алгоритмом RPOP;

– изменив количество скрытых слоев и количество нейронов в каждом слое;

– уменьшив количество входных переменных (в этом случае следует определить новое значение требуемого количества нейронов).

Удалось ли значительно улучшить общую ошибку классификации?

10. Сравните качество моделей между собой и с моделью, построенной на шаге Дерево решений.

11. Сформулируйте рекомендации по использованию результатов моделирования в работе компании.

Контрольное задание:

Предлагается решить задачу из области коллекторского скоринга на тему «Оптимизация выбора стратегии взыскания просроченной задолженности».

В файле debtors.txt имеются статистические данные о заемщиках банка, которые стали должниками, то есть имеют просроченную задолженность. Структура полей файла приведена в следующей таблице.

Имя поля	Описание	Тип данных
Код	Код заемщика	Целый
Сумма кредита	Сумма кредита в рублях	Вещественный
Срок кредита	Срок кредита в мес. (6, 12, 18)	Целый
Ежемесячный платеж	Размер аннуитетного ежемесячного платежа в рублях	Вещественный
Возраст	Возраст, лет	Целый

Пол	Пол клиента	Строковый
Срок просрочки	Срок просрочки в днях	Целый
Число платежей до выхода на просрочку	Сколько сделал заемщик платежей до того, как была впервые зафиксирована просроченная задолженность	Целый
Исполнительный лист	Имеется ли по заемщику исполнительный лист (судебное решение)	0 / 1
Остаток долга (тело)	Сумма в рублях остатка «тела» кредита	Вещественный
Остаток долга (проценты)	Сумма в рублях остатка «процентов» по кредиту	Вещественный
Штрафы	Начисленные штрафы и пенни	Вещественный
Просрочка/Долг	Отношение, рассчитанное на основе трех предыдущих полей по формуле: (12) / (11 + 10)	Вещественный
Возобновление платежей	Начал ли заемщик снова оплачивать кредит после работы специалиста коллекторского подразделения: 0 – работа велась, и возобновления платежей не последовало, 1 – работа велась, и последовало возобновление платежей, «пусто» – работа не велась	Целый
Тестовое множество	Признак участия примера в тестовом множестве	Логический

Известна следующая дополнительная информация:

Ориентировочная стоимость плана мероприятий по взысканию просроченной задолженности с 1 заемщика оценивается в 1 000 рублей (зарплата персоналу, оплата звонков, выездов и т.д.)

Средняя ожидаемая сумма погашения от 1 должника 12 000 рублей.

Задание:

1. Построить модели, позволяющие классифицировать должников на 2 класса в зависимости от исхода мероприятий, проводимого специалистом коллекторского подразделения банка (логистическая регрессия, дерево решений, многослойный персептрон). Разделение на обучающее и тестовое множество уже проведено – используйте поле Тестовое множество. За положительный исход примите Возобновление платежей = 1.

2. Построить LIFT-кривые для моделей (там, где это возможно) и сравнить модели.

3. Используя дополнительную информацию, выбрать лучшую модель с двух позиций:

– Максимизация ожидаемых поступлений от должников;

– Минимизация рабочего времени, которое будет потрачено на

мероприятия по взысканию.

4. Ответьте на вопрос: какой из алгоритмов оказался наименее чувствителен к несбалансированности классов? Комментарии впишите к любому узлу-обработчику.

5. Объедините лучшие модели в Ансамбль моделей.

6. «Прогоните» через ансамбль моделей тех должников, к которым еще не применялись мероприятия по взысканию и получите список клиентов, с которыми нужно работать, ранжированный по убыванию вероятности положительного исхода.

7. Все важные визуализаторы вывести на панель отчетов. Свои комментарии впишите к любому узлу-обработчику.

Тема 5. Пропозиционная логика. ПКс ОС II - 6.1

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Дайте формальную постановку задачи кластеризации
2.	Каковы цели кластеризации?
3.	Что такое метрика расстояния?
4.	Каковы основные шаги алгоритма k-means?
5.	Как в алгоритме k-means определяется принадлежность точки к тому или иномум кластеру?
6.	Что такое центроид (центр тяжести кластера) и какова его роль в алгоритме k-means?
7.	Что является условием остановки алгоритма k-means?
8.	Каковы преимущества и недостатки алгоритма k-means?
9.	Какова структура сети Кохонена? Какова роль входных и выходных нейронов карты?
10.	В чем заключаются процедуры конкуренции, объединения и подстройки весов?
11.	По какому правилу производится подстройка весов нейронов карты?
12.	Что такое радиус обучения и какую роль он играет в процессе построения карты?

13.	Чем определяется требуемое число выходных нейронов сети Кохонена?

Практико-ориентированное задание (компьютерный практикум):

Сегментация клиентов компании

Легенда: руководство отдела маркетинга сети книжных магазинов по продаже книг различной направленности (от деловой литературы до художественной и развлекательной) поставило задачу сегментации своих постоянных покупателей, выявления их предпочтений. Предполагается, что это будет использоваться в будущем для оптимизации ассортимента книжных магазинов, проведения рекламных акций и для других целей.

В качестве исходных данных для сегментации выступают данные о постоянных покупателях книжной сети, у которых имеются дисконтные карты. Благодаря ним, о них известна следующая информация: пол, возраст и осуществленные покупки.

Задание. Вы выступаете в роли бизнес-аналитика. В вашем распоряжении имеются данные клиентах и их покупках за последний год (файл bookstore.txt).

Название поля	**Описание**
Код клиента	Уникальный код клиента
Возраст	Возраст клиента
Пол	Пол клиента
Посещений за год	Число посещений клиентом книжных магазинов, закончившихся покупкой товаров
Книги категории A	Число приобретенных за год клиентом книг из категории A
Книги категории B	Число приобретенных за год клиентом книг из категории B
Книги категории C	Число приобретенных за год клиентом книг из категории C
Книги категории D	Число приобретенных за год клиентом книг из категории D
Книги категории E	Число приобретенных за год клиентом книг из категории E
Среднее число книг в чеке	Среднее число книг в чеке (за год)
Средняя сумма чека	Средняя сумма чека (за год)

Для удобства проведения анализа все книги разделены на пять основных категорий.

Категория книжной продукции	**Описание**
Книги категории A	Деловая, научная, техническая и учебная

	литература
Книги категории B	Художественная литература
Книги категории C	Книги по литературе и искусству и культуре
Книги категории D	Народная медицина и здоровье
Книги категории E	Детективы, любовные романы, фантастика

Перед вами стоит задача сегментации клиентской базы при помощи самоорганизующихся карт Кохонена.

1. Разбейте всех клиентов на сегменты. Перед построением аналитической модели рекомендуется ответить на следующие вопросы:

- Какие признаки включить в алгоритм кластеризации?
- Какие их признаков более значимые, какие – нет?
- Какой размер карты Кохонена выбрать?

2. Проанализировав карты, определите, какое количество кластеров дает самую наглядную картину?

3. Оцените емкость каждого сегмента (мощность кластера).

4. Охарактеризуйте каждый сегмент и дайте ему краткое название.

5. Постарайтесь описать социальный портрет покупателей каждого сегмента. Дайте название каждому сегменту. При интерпретации кластеров вы можете привлекать любую статистическую и (или) графическую информацию по кластеру: количество объектов, среднее, минимум, максимум и т.д.

Тема 6. Анализ и прогнозирование временных рядов. ПКс ОС II - 6.1

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Детерминированная и случайная составляющие временного ряда
2.	Модели временных рядов
3.	Компоненты временного ряда
4.	Исследование временных рядов и автокорреляция
5.	В чем заключается принцип построения ретропрогноза?
6.	Оценка адекватности выбранного метода прогнозирования
7.	Методы сглаживания

8.	Прогнозирование методом скользящего среднего
9.	Декомпозиция временного ряда
10.	Регрессионный анализ временных рядов

Практико-ориентированное задание (компьютерный практикум):

Прогнозирование объема продаж

Легенда: оптовая компания занимается сбытом строительных материалов, ассортимент насчитывает несколько тысяч торговых позиций, объединенных в группы (сухие смеси, плитка, напольные покрытия, грунтовка и т.п.)

Перед вами поставлена задача: автоматизировать ежемесячный расчет потоварного прогноза на следующие три периода.

1. Подключить хранилище данных materials.gdb

2. Для импорта выбрать только процесс Продажи

3. Выбрать для импорта все измерения, атрибуты и факты процесса Продажи, для фактов агрегирующей функцией является Сумма.

4. Определить срез для выбранных измерений. Типовой сценарий будет построен на примере товарной группы «Грунтовка». Установить фильтр «Кроме последнего периода», т.к. данные текущего, еще не завершившегося месяца, в построении модели участия не принимают.

5. Для последующей группировки по месяцам выполнить Преобразование даты к виду «Дата (Год+Месяц)».

6. Выполнить Группировку по измерению «Дата (Год+Месяц)» и факту «Сумма» и/или «Количество».

7. Построить график продаж. Оценить, содержит ли Кривая продаж шумы и выбросы, которые необходимо удалить для получения более качественного прогноза. Для сглаживания данных воспользоваться «Парциальной обработкой».

Парциальная обработка может включать в себя удаление аномалий и сглаживание (фильтрацию) данных (если требуется). При настройке парциальной обработки указать для поля Количество:

- Редактирование аномальных значений с о Степенью подавления (Малая),

- Сглаживание данных с Полосой пропускания — 80.

Отобразить результат преобразования данных на диаграмме. Оценить качество сглаживания.

8. С помощью обработчика «Автокорреляция» определить сезонность продаж данной группы товара. Какие значения АКФ говорят о наличии Тренда и наличии Сезонности?

9. Для построения модели прогноза воспользоваться инструментом «Скользящее окно». Скользящее окно — преобразование временного ряда к таблице смежных значений.

10. Построить несколько моделей временных рядов.

В качестве моделей использовать:

- «Наивную» модель скользящего среднего за предыдущие три месяца.
- Линейную регрессию с периодами (1,2,3) и (1,2,12).
- Нейронную сеть периодами (1,2,12) — Рекомендуется перебрать 2-3 модели нейронной сети, изменяя число нейронов в скрытом слое.

В качестве визуализаторов выбрать Диаграмму рассеивания и Диаграмму, на которой отобразить Количество и Количество_OUT (ретропрогноз).

11. По диаграммам оценить качество моделирования.

12. Вычислить среднюю процентную ошибку для каждой модели. Оценить качество моделирования.

13. К узлу лучшей модели применить обработчик «Прогнозирование». Обработчик Прогнозирование — позволяет получать предсказание значений временного ряда на число отсчетов, соответствующее заданному горизонту прогнозирования.

14. К узлу Прогнозирование применить Фильтр: Шаг прогноза = Не пустой.

15. После получения прогнозных цифр по группе товаров произвести разгруппировку прогнозных значений для получения прогноза для каждой товарной позиции (обработчик «Разгруппировка»).

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ,	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой

	которое может набрать обучающийся		балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ 1	100	0,2	20
КТ 2	100	0,2	20
КТ 3	100	0,2	20
Итого:	х	0,6	60

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ х Коэффициент веса контрольной точки.

5.4. Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ – 1.

Тема 1 и 2.

Опрос:

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Что такое очистка данных?
2.	Что такое предобработка данных?
3.	Что такое обогащение данных?
4.	Что такое редактирование данных?
5.	Что такое кодирование данных?
6.	Какие элементы интерфейса Power BI Desktop предназначены для визуализации данных?
7.	Как работает панель визуализаций?
8.	Как переключаться между режимами представления данных в Power BI Desktop?
9.	Как загрузить данные в Power BI из различных источников (Excel, SQL, веб-сервисы и т. д.)?

10.	Как трансформировать данные в редакторе Power Query?

Критерии оценивания опроса:

Диапазон баллов	Описание критерия
85-100	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
65-84	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.
55-64	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
0-54	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

КТ – 2.

Тема 3 и 4.

Опрос:

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Что такое классификация в статистических методах?
2.	Какие виды группировок используются в статистике?
3.	Как определяется количество групп при группировке по

	атрибутивному признаку?
4.	Как образуется укрупнённая группа если атрибутивный признак имеет множество наименований?
5.	Как рассчитывается количество групп при группировке по количественным признакам?
6.	Что такое классификация в машинном обучении?
7.	Цель классификации
8.	Некоторые методы классификации?
9.	Как модели классификации учатся отделять данные на разные классы?
10.	Какие функции потерь используются в классификации?

Критерии оценивания опроса:

Диапазон баллов	Описание критерия
85-100	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
65-84	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.
55-64	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
0-54	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

КТ – 3.

Тема 5 и 6.
Опрос:

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	В чём состоит задача кластерного анализа?
2.	К какой предметной области может быть применён кластерный анализ?
3.	В чём заключается практическая ценность кластерного анализа?
4.	Как классифицируются задачи кластерного анализа с точки зрения информации о числе кластеров?
5.	Что является результатом кластерного анализа?
6.	Что такое временной ряд?
7.	Какие бывают временные ряды?
8.	Что такое стационарный и нестационарный временной ряд?
9.	Назовите составляющие временного ряда.
10.	Что собой представляет тренд?

Критерии оценивания опроса:

Диапазон баллов	Описание критерия
85-100	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
65-84	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.

55-64	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
0-54	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

5.5. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Обучающемуся разрешается использование компьютера с установленными прикладными программными продуктами

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация проводится в форме зачета с оценкой.

6.2. Типовые оценочные материалы промежуточной аттестации

Типовые проверочные задания для самоподготовки обучающегося к промежуточной аттестации:

Тема 1. Подготовка данных к анализу. ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Аналитический и информационный подход к моделированию
2.	Основные этапы построения моделей
3.	Формы представления, типы и виды данных
4.	Классификация задач Data Mining
5.	Виды источников данных

6.	Хранилища данных
7.	Многомерная модель данных
8.	Оценка качества данных
9.	Предобработка данных
10.	Фильтрация данных
11.	Обработка дубликатов и противоречий
12.	Выявление аномальных значений
13.	Восстановление пропущенных значений
14.	Сокращение размерности
15.	Трансформация данных
16.	Группировка данных
17.	Слияние данных
18.	Квантование
19.	Нормализация и кодирование

2. Задания комбинированного типа:

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Процесс обнаружения и исправления (или удаления) записей с пропущенными, некорректными или противоречивыми значениями называется _____ данных. Варианты ответов: а) очисткой б) агрегацией		
2.	Значения, которые существенно отличаются от остальных наблюдений в наборе данных, называются _____. Варианты ответов: а) выбросами		

	б) пропусками		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только букву выбранного варианта ответа (а)).

Какой этап работы с данными обычно предшествует анализу и включает очистку, преобразование и организацию данных?

а) Визуализация данных;

б) Построение моделей машинного обучения;

в) Предобработка (подготовка) данных;

г) Формулировка выводов.

Что такое «пропущенные значения» в наборе данных?

а) Значения, равные нулю;

б) Отсутствующие данные в некоторых полях записей;

в) Аномально большие значения;

г) Некорректные форматы данных.

Какой из перечисленных методов НЕ является стандартным способом обработки пропущенных значений?

а) Удаление строк или столбцов с пропусками;

б) Заполнение пропусков средним, медианой или модой;

в) Использование моделей для прогнозирования пропущенных значений;

г) Замена всех пропусков максимальным значением в столбце.

Что такое «выбросы» в данных?

а) Наиболее часто встречающиеся значения;

б) Наблюдения, которые значительно отклоняются от остальных данных и могут быть вызваны ошибками измерения или редкими событиями;

в) Пропущенные значения;

г) Столбцы с текстовыми данными.

Тема 2. Визуализация данных. ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Цели и задачи визуализации данных в аналитических технологиях
2.	Виды визуализаторов данных
3.	OLAP-анализ

1.2. Контрольные задания.

Задание 1.

Имеются данные о продажах книжного магазина за первое полугодие 2024 года по четырём категориям книг (в штуках):

Месяц	Художественная	Научная	Детская	Учебная
Январь	120	45	200	150
Февраль	115	50	180	160
Март	130	55	220	140
Апрель	125	60	210	155
Май	140	65	230	145
Июнь	135	70	240	135

Постройте четыре различных типа графиков на основе предоставленных данных:

1. Столбчатая диаграмма — показывающая общий объём продаж по месяцам (сумма всех категорий).

2. Линейный график — отображающий динамику продаж по каждой категории за все 6 месяцев.

3. Круговая диаграмма — демонстрирующая долю каждой категории в общих продажах за весь период.

4. Тепловая карта — где по оси X — категории книг, по оси Y — месяцы, а цвет ячейки отражает объём продаж.

2. Задания комбинированного типа.

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Основная цель визуализации данных — _____ сложную информацию в графической форме для облегчения восприятия, анализа и коммуникации. Варианты ответов: а) представить б) скрыть		
2.	Гистограмма используется для визуализации _____ одной числовой переменной. Варианты ответов: а) распределения б) корреляции		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только букву выбранного варианта ответа (а)).

Для визуализации взаимосвязи между двумя непрерывными переменными обычно используется:

а) Столбчатая диаграмма;

б) Диаграмма рассеяния;

в) Круговая диаграмма;

г) Линейчатая диаграмма.

Тест 2.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать несколько правильных ответов.

4. Записать только буквы выбранного варианта ответа (например, а), в)).

Какие из перечисленных принципов являются важными для создания эффективной визуализации данных?

а) Минимизация «визуального шума» (удаление лишних линий, теней, украшений);

б) Использование осмысленных подписей, заголовков и легенд;

в) Обязательное использование 3D-эффектов для привлекательности;

г) Выбор типа графика, соответствующего типу данных и цели;

д) Одновременное отображение максимального количества переменных на одном графике.

Какие типы графиков подходят для отображения изменения величины во времени (временного ряда)?

а) Круговая диаграмма;

б) Линейный график;

в) График с областями;

г) Диаграмма рассеяния.

Тест 3. Установите соответствие между типом данных/задачей, приводимом в столбце слева, и рекомендуемым типом визуализации из правого столбца: к каждой позиции, данной в левом столбце, подберите соответствующую позицию из правого столбца.

1	Сравнение долей целого	А	Круговая диаграмма
2	Показать	Б	Тепловая карта

	распределение одной категориальной переменной		
3	Отобразить корреляцию между многими числовыми переменными	В	Диаграмма дерева
4	Визуализировать иерархические данные или вложенность	Г	Столбчатая диаграмма

1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов.

2. Внимательно прочитать оба списка:

список 1 – вопросы, утверждения, факты, понятия и т.д.;

список 2 – утверждения, свойства объектов и т.д.

3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов.

4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа в таблицу (например, 1/А или 4/Б).

1	2	3	4

Тема 3. Классификация и регрессия (Статистические методы). ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Задача классификации
2.	Статистические методы решения задачи классификации

1.2. Контрольные задания.

Задание 1.

Даны данные о зависимости времени подготовки к экзамену (в часах) и

полученной оценки (по 100-балльной шкале):

Студент	Часы (X)	Оценка (Y)
1	5	45
2	10	60
3	15	75
4	20	85
5	25	95

1. Рассчитайте коэффициенты линейной регрессии $Y = b_0 + b_1X$ методом наименьших квадратов.

2. Сделайте прогноз оценки для студента, который готовился 18 часов.

3. Рассчитайте коэффициент детерминации R^2 и интерпретируйте его.

Задание 2.

Исследуется зависимость сдачи экзамена (1 — сдал, 0 — не сдал) от количества часов подготовки (X).

Известно, что логистическая функция имеет вид:

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1X)}}$$

где $b_0 = -2{,}5, b_1 = 0{,}3$

2. Задания комбинированного типа.

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Задача _____ предсказывает категориальную целевую переменную (например, класс объекта). Варианты ответов: а) классификации б) регрессии		
2.	Задача _____ предсказывает непрерывную числовую целевую переменную		

	(например, цену или рост). Варианты ответов: а) регрессии б) кластеризации		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только букву выбранного варианта ответа (а)).

Что является результатом работы модели логистической регрессии для одного объекта?

а) Конкретный класс;

б) Вероятность принадлежности объекта к целевому классу;

в) Вещественное число, которое может быть любым;

г) Расстояние до центра кластера.

Как называется простая регрессионная модель, описывающая линейную зависимость одной зависимой переменной от одной независимой?

а) Множественная линейная регрессия;

б) Парная линейная регрессия;

в) Полиномиальная регрессия;

г) Логистическая регрессия.

Тест 2.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать несколько правильных ответов.

4. Записать только буквы выбранного варианта ответа (например, а), в)).

Какие из следующих утверждений о методе k-ближайших соседей для классификации верны?

а) Это непараметрический метод: модель не делает предположений о виде зависимости между признаками и целевой переменной;

б) На этапе обучения он просто запоминает все обучающие данные;

в) Основная вычислительная нагрузка приходится на этап обучения (построение модели);

г) При прогнозе для нового объекта алгоритм находит k наиболее похожих (близких) объектов в обучающей выборке и относит его к классу, который преобладает среди этих соседей.

Какие из перечисленных метрик используются для оценки качества моделей регрессии?

а) Среднеквадратическая ошибка (MSE);

б) Accuracy (Точность);

в) F1-score;

г) Коэффициент детерминации R^2.

Тема 4. Классификация и регрессия (Машинное обучение). ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Деревья решений
2.	Нейронные сети
3.	Эффект переобучения
4.	Оценка качества классификационных моделей

1.2. Контрольные задания.

Задание 1.

Даны данные о погоде и решении пойти на прогулку:

День	Температура	Влажность	Ветер	Прогулка
1	Жарко	Высокая	Нет	Нет
2	Жарко	Высокая	Да	Нет
3	Тепло	Высокая	Нет	Нет
4	Прохладно	Нормальная	Нет	Да
5	Прохладно	Нормальная	Да	Нет
6	Прохладно	Нормальная	Да	Да

1. Постройте дерево решений для классификации (идти на прогулку или нет).

2. Какой признак будет выбран корневым узлом? Объясните, почему.

3. Сделайте прогноз для нового дня: Температура=Тепло, Влажность=Нормальная, Ветер=Нет.

2. Задания комбинированного типа.

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Алгоритм, который строит дерево решений, разбивая данные на подмножества... Варианты ответов: а) Дерево решений б) Случайный лес		
2.	Ансамблевый метод _____ объединяет множество деревьев решений, обучаемых на случайных подвыборках данных и признаков, что снижает переобучение и повышает точность. Варианты ответов: а) Random Forest (случайный лес)		

	б) Gradient Boosting		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только последовательность букв выбранного варианта ответа (а)).

Как называется фундаментальное разделение данных на две части: одну для обучения модели, а другую — для объективной оценки её качества?

а) Кластеризация;

б) Разделение на обучающую и тестовую выборки (train-test split);

в) Балансировка классов;

г) Кросс-валидация.

Какой алгоритм машинного обучения для классификации и регрессии строит иерархическую структуру решающих правил типа «если-то»?

а) Метод опорных векторов (SVM);

б) Дерево решений (Decision Tree);

в) Логистическая регрессия;

г) Метод k-ближайших соседей.

Что такое «ансамблевые методы» (ensemble methods) в машинном обучении?

а) Использование только одной, но очень сложной модели;

б) Комбинирование прогнозов нескольких базовых моделей (слабых учеников) для получения более точного и устойчивого результата;

в) Обучение модели исключительно на новых данных;

г) Методы, исключающие необходимость в тестовых данных.

Тест 2.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать несколько правильных ответов.

4. Записать только буквы выбранного варианта ответа (например, а), в)).

Какие из следующих алгоритмов относятся к ансамблевым методам?

а) Случайный лес (Random Forest);

б) Линейная регрессия;

в) Градиентный бустинг (Gradient Boosting);

г) Метод k-средних (k-means);

д) Наивный байесовский классификатор.

Тема 5. Кластеризация. ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Задача кластеризации
2.	Алгоритмы решения задачи кластеризации
3.	Алгоритм k-means
4.	Сети и карты Кохонена

2. Задания комбинированного типа.

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Кластеризация — это задача _____, целью которой является группировка объектов так, чтобы объекты внутри одной группы		

	(кластера) были более похожи друг на друга, чем на объекты из других групп. Варианты ответов: а) обучения без учителя (unsupervised learning) б) обучения с учителем (supervised learning)		
2.	Алгоритм _____ — один из самых известных методов кластеризации, который итеративно назначает объекты ближайшему центроиду и пересчитывает центроиды как среднее точек кластера. Варианты ответов: а) K-means (K-средних) б) DBSCAN		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только букву выбранного варианта ответа (а)).

Кластеризация — это задача машинного обучения без учителя, целью которой является:

а) Предсказание целевой переменной по историческим данным;

б) Разбиение множества объектов на группы (кластеры) таким образом, чтобы объекты внутри одного кластера были похожи друг на друга, а

объекты из разных кластеров — отличались;

в) Обнаружение аномалий в данных путём сравнения с заданным шаблоном;

г) Понижение размерности данных.

Какой алгоритм кластеризации способен находить кластеры произвольной формы и автоматически определять выбросы (шум), не требуя задания числа кластеров?

а) k-средних;

б) Иерархическая кластеризация;

в) DBSCAN (Density-Based Spatial Clustering of Applications with Noise);

г) Алгоритм сдвига среднего (Mean Shift).

Что является результатом работы алгоритма иерархической кластеризации?

а) Набор независимых кластеров с центрами;

б) Дендрограмма — древовидная структура, показывающая последовательность слияния или разделения кластеров на всех уровнях;

в) Список меток кластеров без какой-либо иерархии;

г) График зависимости качества от числа кластеров.

Что такое «проклятие размерности» в контексте кластеризации и многих других задач машинного обучения?

а) Слишком малое количество признаков;

б) В пространствах высокой размерности все точки становятся практически равноудалёнными друг от друга, что делает понятие близости и плотности менее осмысленным, а алгоритмы — неэффективными;

в) Невозможность визуализировать данные в 3D;

г) Автоматическое увеличение числа кластеров.

Тема 6. Анализ и прогнозирование временных рядов. ПКс ОС II - 6.1

1. Задания открытого типа.

1.1. Вопросы открытого типа.

Прочитайте текст и запишите развернутый обоснованный ответ:

№ п.п.	Вопрос
1.	Модели временных рядов
2.	Методы сглаживания временных рядов
3.	Выбор метода прогнозирования
4.	Оценка адекватности выбранного метода прогнозирования

2. Задания комбинированного типа.

2.1. Тестовые задания с обоснованием выбора.

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1.	Временной ряд — это последовательность данных, упорядоченных по _____, часто с равными интервалами. Варианты ответов: а) времени б) категориям		
2.	_____ — это долгосрочное направленное движение данных (рост, спад или стагнация) во временном ряде. Варианты ответов: а) Тренд (Trend) б) Сезонность (Seasonality)		

3. Задания закрытого типа.

3.1. Тестовые задания.

Тест 1.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только букву выбранного варианта ответа (а)).

Какой метод прогнозирования основан на среднем значении ряда за предыдущие периоды, где более свежим наблюдениям придаются большие веса?

а) Наивный прогноз;

б) Метод экспоненциального сглаживания;

в) Метод скользящего среднего;

г) ARIMA-моделирование.

Для чего используется критерий Дики-Фуллера (ADF test) в анализе временных рядов?

а) Для проверки наличия гетероскедастичности остатков;

б) Для проверки автокорреляции остатков;

в) Для проверки гипотезы о наличии единичного корня (тест на стационарность);

г) Для выявления сезонности.

Тест 2.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов.

2. Внимательно прочитать предложенные варианты ответа.

3. Выбрать несколько правильных ответов.

4. Записать только буквы выбранного варианта ответа (например, а), в)).

Какие из перечисленных методов являются простыми (наивными) методами прогнозирования временных рядов?

а) Прогноз как последнее известное значение (Last Value);

б) Прогноз как среднее значение всех прошлых наблюдений (Mean);

в) Прогноз с помощью модели SARIMA;

г) Прогноз с использованием нейронных сетей (LSTM).

Какие компоненты обычно выделяют в классической аддитивной или мультипликативной декомпозиции временного ряда?

а) Тренд (T);
б) Сезонность (S);
в) Тренд-сезонность (TS);
г) Остаток/Случайная компонента (R);
д) Трансформация (L).

6.3. Критерии и шкала оценивания на основе БРС.

КРИТЕРИИ ОЦЕНИВАНИЯ	РЕЗУЛЬТАТ В БАЛЛАХ
Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок	40
Дан развернутый ответ на поставленный вопрос, где обучающийся демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры, в ответе присутствует свободное владение монологической речью, логичность и последовательность ответа. Однако допускается неточность в ответе. Решил предложенные практические задания с небольшими неточностями.	30-39
Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры, недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа и решении практических заданий.	20-29
Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Решение практических заданий не выполнено, т.е. обучающийся не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.	0-19

6.4. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий.

Обучающемуся разрешается использование компьютера с установленными прикладными программными продуктами

7. Методические материалы по освоению дисциплины (модуля)

Подготовка к лекциям.

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому обучающемуся следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции.

Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность обучающегося. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим обучающимся. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось просить их у однокурсников и тем самым не отвлекать их во

время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Подготовка к практическим занятиям.

Подготовку к каждому практическому занятию каждый обучающийся должен начать с ознакомления с планом практического занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений обучающемуся необходимо самостоятельно выбрать тему доклада по проблеме практического занятия и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или 10 письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности обучающегося свободно ответить на теоретические вопросы практического занятия, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Структура практического занятия:

В зависимости от содержания и количества отведенного времени на изучение каждой темы может практическое занятие состоять из четырех-пяти частей:

1. Обсуждение теоретических вопросов, определенных программой дисциплины.
2. Доклад и/ или выступление с презентациями по проблеме практического занятия.
3. Обсуждение выступлений по теме – дискуссия.
4. Выполнение практического задания с последующим разбором полученных результатов или обсуждение практического задания, выполненного дома, если это предусмотрено программой.
5. Подведение итогов занятия.

Первая часть – обсуждение теоретических вопросов - проводится в виде фронтальной беседы со всей группой и включает выборочную проверку преподавателем теоретических знаний обучающихся. Примерная продолжительность — до 15 минут. Вторая часть — выступление обучающихся с докладами, которые должны сопровождаться презентациями с целью

усиления наглядности восприятия, по одному из вопросов практического занятия. Обязательный элемент доклада – представление и анализ статистических данных, обоснование социальных последствий любого экономического факта, явления или процесса. Примерная продолжительность — 20-25 минут. После докладов следует их обсуждение – дискуссия. В ходе этого этапа практического занятия могут быть заданы уточняющие вопросы к докладчикам. Примерная продолжительность – до 15-20 минут. Если программой предусмотрено выполнение практического задания в рамках конкретной темы, то преподавателями определяется его содержание и дается время на его выполнение, а затем идет обсуждение результатов. Если практическое задание должно было быть выполнено дома, то на практическом занятии преподаватель проверяет его выполнение (устно или письменно). Примерная продолжительность – 15-20 минут. Подведением итогов заканчивается практическое занятие. Обучающимся должны быть объявлены оценки за работу и даны их четкие обоснования. Примерная продолжительность — 5 минут.

Работа с литературными источниками.

В процессе подготовки к практическим занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем, что позволяет обучающимся проявить свою индивидуальность в рамках выступления на занятиях, выявить широкий спектр мнений по изучаемой проблеме.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Полковникова, Н. А. Анализ и визуализация данных в Microsoft Excel в примерах и задачах : практическое пособие / Н. А. Полковникова. - Москва ; Вологда : Инфра-Инженерия, 2023. - 172 с. - ISBN 978-5-9729-1485-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2092453

2. Гинько, А. Ю. Анализ и визуализация данных в Yandex DataLens.

Подробное руководство: от новичка до эксперта : практическое руководство / А. Ю. Гинько. - Москва : ДМК Пресс, 2023. - 358 с. - ISBN 978-5-93700-171-9. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2155919

8.2. Дополнительная литература

1. Чубукова, И. А. Data Mining: учебное пособие / И. А. Чубукова. — 3-е изд. — Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 469 с. — ISBN 978-5-4497-0289-0. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: http://www.iprbookshop.ru/89404.html

2. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москв : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/450262

3. Подкорытова, О. А. Анализ временных рядов: учебное пособие для вузов / О. А. Подкорытова, М. В. Соколов. — 2-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2020. — 267 с. — (Высшее образование). — ISBN 978-5-534-02556-9. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/450587

8.3. Нормативные правовые документы и иная правовая информация

1. Федеральный закон Российской Федерации от 27 июля 2006 г. N 149-ФЗ Об информации, информационных технологиях и о защите информации, Собрание Законодательства РФ. Официальное издание. 2006 г.

2. Федеральный закон "Об информации, информатизации и защите информации", от 20 февраля 1995 Г. N 24-ФЗ (с изменениями от 10 января 2003 г.)

8.4. Интернет-ресурсы

1. www.biblio-online.ru — Электронно-библиотечная система [ЭБС] Юрайт;
2. http://www.iprbookshop.ru — Электронно-библиотечная система [ЭБС] «Iprbooks»
3. https://e.lanbook.com — Электронно-библиотечная система [ЭБС] «Лань».
4. https://dlib.eastview.com — East View Information Services, Inc. Коллекции электронных научных и практических журналов.
5. http://elibrary.ru/ — Научная электронная библиотека Elibrary.ru.
6. http://www.intuit.ru — Национальный Открытый Университет «ИНТУИТ»;

7. http://www.planetaexcel.ru — сайт о возможностях MS Excel (приемы, видео-уроки, книги);

8. http://www.clouderp.ru/tags/BUSINESS_INTELLIGENCE/ Что такое Business Intelligence? Обзор BI систем

9. http://netpeak.net/ru/blog/kak-rabotat-s-microsoft-power-bi-podrobnoe-rukovodstvo/ Как работать с Microsoft Power BI — подробное руководство

10. http://community.powerbi.com/ Форум (англ. яз.) по Microsoft Power BI. Примеры применения. Ответы на вопросы. Использование на мобильных устройствах и др.

11. https://www.microsoft.com/ru-RU/download/details.aspx?id=45331 – ссылка для скачивания Microsoft Power BI;

12. https://docs.microsoft.com/ru-ru/power-bi/desktop-getting-started – техническая документация Microsoft, документация Power BI;

13. https://support.office.com/ru-ru/article/Вход-в-sway-db30c63b-6548-456d-b0e29fcd68bfa7a0?ui=ru-RU&rs=ru-RU&ad=RU – руководство по созданию презентации в MS Sway;

14. http://www.gks.ru – официальный сайт Федеральной службы государственной статистики;

15. http://moscow.gks.ru – Территориальной орган Федеральной службы государственной статистики по городу Москве;

16. https://data.mos.ru – портал открытых данных Правительства Москвы

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

9.1. Материально-техническая база

Перечень материально-технического обеспечения:

1. Учебные аудитории, оборудованные для проведения учебных занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, коллоквиумов, мероприятий текущего контроля и промежуточной аттестации, в том числе мультимедийным оборудованием для демонстрации электронных презентаций и аудио- и видеоматериалов;

2. Компьютерные классы для выполнения групповых тестовых и иных заданий, а также для самостоятельной работы обучающихся оснащенные компьютерной техникой и обеспечением доступа к сети «Интернет» и доступа в электронную информационно-образовательную среду организации;

3. Специализированные аудитории и лаборатории;

4. Библиотека с обеспечением печатными изданиями или электронно-библиотечная система обеспечивающая доступ к электронным изданиям (электронная библиотека);

5. Читальный зал;

6. Технические средства обучения: персональные компьютеры; компьютерные проекторы; звуковые динамики; программные средства,

обеспечивающие просмотр видеофайлов в форматах AVI, MPEG-4, DivX, RMVB, WMV и др.

9.2.Информационные технологии, программное обеспечение:

При осуществлении образовательного процесса по дисциплине используется информационные технологии и программное обеспечение:

1. Современная операционная система.
2. Kaspersky Endpoint Security (или аналог).
3. Средство просмотра файлов формата pdf.
4. Современные офисные средства (текстовые и табличные редакторы, средства работы с презентационными материалами и т.д.).
5. Архиватор 7-Zip.
6. Система дистанционного обучения.
7. Автоматизированная библиотечная система.
8. СУБД PostgreSQL.
9. Lucidchart.
10. Eclipse
11. Система ADVANTA для управления проектами.

9.3. Информационные справочные системы:

1. https://www.iprbookshop.ru – Электронно-библиотечная система [ЭБС] «IPR SMART»
2. https://e.lanbook.com - Электронно-библиотечная система [ЭБС] «Лань».
3. https://znanium.ru - Электронно-библиотечная система [ЭБС] «Znanium.com».
4. https://www.ibooks.ru/ - Электронно-библиотечная система [ЭБС] «ibooks.ru»
5. http://www.consultant.ru/ - Справочно-правовая система «Консультант».
6. http://www.garant.ru/ - Справочно-правовая система «Гарант».