

Федеральное государственное бюджетное учреждение науки Институт проблем региональной экономики Российской академии наук

Утверждена на заседании Ученого совета ФГБУН Институт проблем региональной экономики Российской академии наук протокол №8 от «19» июня 2023 г.

«УТВЕРЖДАЮ»

Директор ФГБУН Институт проблем региональной экономики Российской академии наук



Д.э.н., проф. проф. РАО А.Д. Шматко
«19» июня 2023 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
реализуемой без применения электронного (онлайн) курса

Б1.В.ДВ.04.01 Анализ данных

(код и наименование РПП)

Направление подготовки: 38.04.01 «Экономика»

Профиль: Региональная экономика

Форма обучения: очная/заочная

Год набора: 2024

Санкт-Петербург 2023 год

Автор–составитель:

ст.н.с. лаборатории комплексного исследования пространственного развития регионов, к.э.н. Е.А. Назарова

ст.н.с. лаборатории комплексного исследования пространственного развития регионов, к.э.н. Л.В. Дорофеева

Содержание

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.....	4
2. Объем и место дисциплины в структуре ОП.....	4
3. Содержание и структура дисциплины.....	5
4. Материалы текущего контроля успеваемости обучающихся.....	6
5. Оценочные материалы промежуточной аттестации по дисциплине.....	7
6. Методические материалы для освоения дисциплины.....	11
7. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет".....	13
7.1. Основная литература.....	13
7.2. Дополнительная литература.....	13
7.3. Нормативные правовые документы и иная правовая информация.....	13
7.4. Интернет-ресурсы.....	13
7.5. Иные источники.....	13
8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы.....	14

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

1.1 Дисциплина Б1.В.ДВ.04.01 «Анализ данных» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ОПК-3	Способен обобщать и критически оценивать научные исследования в экономике	ОПК-3.1	Способен применять навыки обобщения и структурирования материала научных исследований в экономике
		ОПК-3.2	Способен использовать методы критической оценки научных исследований в экономике в сфере профессиональной деятельности

1.2 В результате освоения дисциплины у студентов должны быть сформированы

Код компонента компетенции	Результаты обучения
ОПК-3.1 ОПК-3.2	уровне знаний: - о актуальных направлениях научных исследований в области оценочных процедур
	на уровне умений: - умением выбрать актуальную тему научного исследования в области оценочной деятельности
	на уровне навыков: - практическими навыками обоснования актуальности выбранной темы научного исследования

2. Объем и место дисциплины в структуре ОП

Общая трудоемкость дисциплины составляет 4 зачетных единиц, 144 академических часов.

Вид работы	Трудоемкость очно/заочно (в акад. часах)
Общая трудоемкость	144/144
Контактная работа с преподавателем	66/14
Лекции	30/8
Практические занятия	30/2
Лабораторные занятия	4/2
Самостоятельная работа	32/115
Консультации	2

Контроль	36/9
Формы текущего контроля	Опрос/тестирование
Форма промежуточной аттестации	экзамен/курсовая работа

Место дисциплины в структуре ОП ВО

Учебная дисциплина Б1.В.ДВ.04.01 «Анализ данных» относится к обязательным дисциплинам базовой части направления подготовки (уровень магистратуры) 38.04.01 «Экономика».

Изучение дисциплины происходит одновременно с изучением таких дисциплин как _____, что обеспечивает успешное освоение профессиональных компетенций. Знания, умения и навыки, полученные при изучении дисциплины, используются магистрантами при выполнении выпускных квалификационных работ (магистерских диссертаций), а также в дальнейшей практической деятельности.

Дисциплина Б1.В.ДВ.04.01 «Анализ данных» изучается:

- магистрантами очной формы обучения на 3 курсе 3 и 4 семестрах. Форма промежуточной аттестации – экзамен.
- магистрантами заочной формы обучения на 2 и 3 курсах. Форма промежуточной аттестации – экзамен.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем,	Объем дисциплины, час.					СР	Форма текущего контроля успеваемости **, промежуточной аттестации** *	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий						
			Л	ЛР	ПЗ	КСР			
Тема 1	Основы анализа данных. Системы поддержки принятия решений	30	10	-	10		10	Т, КР	
Тема 2	Предобработка и очистка данных	33	10	2	10		11	Т, КР	
Тема 3	Классификационный анализ без обучения. Кластерный анализ	33	10	2	10		11	Т, КР	
Промежуточная аттестация		36							Экзамен Кр
Всего:		144	30	4	30	2*	32		

Заочная форма обучения

№ п/п	Наименование тем,	Объем дисциплины, час.					CP	Форма текущего контроля успеваемости **, промежуточной аттестации** *
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					
			Л	ЛР	ПЗ	КСР		
Тема 1	Основы анализа данных. Системы поддержки принятия решений	41	2	-	4		35	Т, КР
Тема 2	Предобработка и очистка данных	46	2	2	2		40	Т, КР
Тема 3	Классификационный анализ без обучения. Кластерный анализ	46	4	-	2		40	Т, КР
Промежуточная аттестация		9						Экзамен Кр
Всего:		144	8	2	8	2*	115	

УО – устный опрос, К – кейс-задание, CP – задания для самостоятельной работы Т – тестирование

3.2. Содержание дисциплины

Тема 1. Основы анализа данных. Системы поддержки принятия решений

Введение. Понятие анализа данных. Задачи систем поддержки принятия решений. OLTP и OLAP-системы. Принципы построения информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Правила Кодда. Размерностные модели. MOLAP, ROLAP, HOLAP- системы. Витрины данных. ETL (Extracting Transforming and Loading) – средство извлечения, обработки и загрузки данных. Добыча данных. Добыча данных в управлении качеством. DataMining. Стандарты DataMining. Стандарт CWM, CRISP, PMML. Жизненный цикл процесса анализа данных. Классификация методов DataMining. Модели DataMining. Понятие данные и знания. Процесс обнаружения знаний. Классификация задач DataMining. Методы анализа данных. Разведочный анализ данных. Очистка и фильтрация данных. Статистические диаграммы. «Ящичные» диаграммы. Диаграммы «ствол-листья». Задачи классификации и регрессии. Использование статистических пакетов для интеллектуального анализа данных. Понятие бизнес-аналитики. Средства бизнес-аналитики. Средства легкой бизнес-аналитики. QlikView, QlikSense.

Тема 2. Предобработка и очистка данных

Методология KDD. Задачи предобработки данных. Технология ETL. Просмотр данных. Очистка данных. Оценка качества данных. Заполнение пропущенных данных. Аномальные и предельные данные. Использование ящечной диаграммы. Выявление дубликатов и противоречий. Корреляционный анализ. Использование факторного анализа

при предобработке данных. Трансформация данных. Квантование. Сэмплинг. Группировка данных.

Тема 3. Классификационный анализ без обучения. Кластерный анализ

Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендограммы. Метод К-средних. Профили кластеров. Взаимосвязь кластерного и регрессионного анализа. Использование пакета Deductor для решения задач кластерного анализа. Кластерный анализ в средствах интеллектуального анализа MicrosoftOffice.

4. Материалы текущего контроля успеваемости обучающихся

4.1. В ходе реализации дисциплины Б1.В.ДВ.04.01 «Анализ данных» используются следующие методы текущего контроля успеваемости обучающихся:

- при проведении занятий лекционного типа: устный опрос;
- при проведении занятий семинарского типа: контрольная работа, тест

Тема и/или раздел	Методы текущего контроля успеваемости* очно/ заочно
Тема 1. Основы анализа данных. Системы поддержки принятия решений	Т, КР
Тема 2. Предобработка и очистка данных	Т, КР
Тема 3. Классификационный анализ без обучения. Кластерный анализ	Т, КР

4.2. Типовые материалы текущего контроля успеваемости обучающихся

4.2.1. Типовые оценочные материалы по темам 1-3

1. Типовые тестовые вопросы.

ЗАДАНИЕ № 1 (выберите несколько вариантов ответа)

Метод главных компонент ...

ВАРИАНТЫ ОТВЕТОВ:

- | | |
|---|---|
| 1) Является частным случаем метода факторного анализа | 2) Предназначен для снижения размерности задачи |
| 3) Устраняет проблему коррелированности факторов | 4) Предназначен для классификации |

ЗАДАНИЕ № 2 (выберите один вариант ответа)

Сигмоидальная активизационная функция искусственного нейрона имеет вид...

4.2.3. Типовые материалы текущего контроля успеваемости обучающихся к теме 3.

Классификационный анализ без обучения. Кластерный анализ

Тестовые задания:

ЗАДАНИЕ № 1 (выберите несколько вариантов ответа)

В кластерном анализе используются методы объединения ...

ВАРИАНТЫ ОТВЕТОВ:

- 1) Ближнего соседа
- 2) Дальнего соседа
- 3) Среднего соседа
- 4) центроидный метод

ЗАДАНИЕ № 2 (выберите несколько вариантов ответа)

В кластерном анализе для определения близости между кластерами используются метрики ...

ВАРИАНТЫ ОТВЕТОВ:

- 1) Эвклидово расстояние
- 2) Куб Эвклидова расстояния
- 3) Взвешенное эвклидово расстояние
- 4) Квадрат Эвклидова расстояния

2. Типовые контрольные работы

Контрольная работа включает пять задач. Шаблоны контрольной работы размещены в файле Excel. К тематике задач относятся: задача очистки данных, иерархическая задача кластерного анализа, решение задачи кластерного анализа методов к-средних, построение ассоциативных правил, построение дерева решений.

Пример задачи. Построить дендограмму, используя Евклидово расстояние и метод "дальнего соседа". Перед построением кластеров выполнить стандартизацию значений атрибутов

Номер объекта	x1	x2
1	3,00	10,00
2	4,00	11,00
3	6,00	10,00
4	10,00	9,00
5	11,00	9,00
6	10,00	7,00

Найти ассоциативные правила, если множества транзакций имеют вид

TID	Предметные наборы			
TID1	зубная паста	крем для бритья	шампунь	
TID2	мыло	дезодорант	шампунь	
TID3	шампунь	дезодорант	лосьен после бритья	шампунь
TID4	крем для бритья	шампунь	дезодорант	лосьен после бритья
TID5	лосьен после бритья	мыло	зубная паста	
TID6	дезодорант	мыло	лосьен после	дезодорант

			бритья	
TID7	дезодорант	шампунь		
TID8	зубная паста	дезодорант	крем для бритья	
TID9	дезодорант	мыло	лосьен после бритья	
TID10	лосьен после бритья	шампунь		

Расчетно-графическое задание (пример).

Использование пакета QlikView для решения задач анализа данных о демографической ситуации в России. Для каждого варианта приведены таблицы с указанием вида исходных данных, которые будут анализироваться средствами бизнес-аналитики

5. Оценочные материалы промежуточной аттестации по дисциплине

5.1. Экзамен проводится с применением следующих методов (средств): в форме решения задач и ответа на вопрос. На зачет выносится материал в объеме, предусмотренном рабочей программой учебной дисциплины.

5.2. Типовые оценочные материалы промежуточной аттестации

Примеры задач

1. Создать файл QlikView с данными о Российских банках. Создать списки, содержащие данные о городе, банке, бюджете, количестве рабочих и активах. Построить гистограмму количества рабочих в банке.
2. Найти описательную статистику и оценить характеристики вариационного ряда. Построить гистограмму распределения для данных на листе Описательная статистика в файле Excel с заданиями на зачет.
3. Создать файл QlikView с данными о Российских банках. Создать списки, содержащие данные о городе, банке, бюджете, количестве рабочих и активах. Создать объект «диаграмма» с измерением «город», содержащий данные о минимальном, максимальном и среднем числе рабочих в банке.
4. Создать файл QlikView с данными о Российских банках. Создать списки, содержащие данные о городе, банке, бюджете, количестве рабочих и активах. Создать объект «диаграмма» с измерением «банк», содержащий данные об активах банков. Отсортировать данные по убыванию. Выбрать 10 банков с максимальными активами. Найти статистические характеристики
5. Выполнить предобработку данных и построить диаграмму «ящик с усами», если в качестве границ ящика использовать – первый и третий квартили. Для определения аномальных значений использовать $1,5 \times [\text{межквартильное расстояние}]$. Для использования крайних значений использовать $2 \times [\text{межквартильное расстояние}]$. Усы строить после удаления аномалий.
6. Создать файл QlikView с данными о нагрузке. Построить гистограмму суммарной нагрузки в группах, используя лист Дисциплины, группы, преподаватели. Построить сводную таблицу с размерностями Группа, преподаватели, фактом – суммарное число часов. Отобразить схему данных с помощью обозревателя таблиц.

Перечень вопросов для экзамена

1. Системы поддержки принятия решений. Хранилища данных.
2. Размерностные модели. OLAP-куб. Таблица размерностей. Таблица фактов.
3. Сравнительный анализ OLAP и OLTP-систем.
4. Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
5. Этапы анализа данных. KDD.
6. Data Mining. Средства обработки Data Mining
7. Элементы математической статистики. Описательная статистика. Операции агрегирования данных.
8. Графические средства анализа. Диаграмма рассеяния. Гистограмма.
9. Начальные этапы KDD. ETL. Средства очистки и трансформации данных.
10. Классификация методов предобработки и очистки данных.
11. Методы борьбы с аномалиями. Ящечная диаграмма.
12. Общая характеристика задач кластерного анализа.
13. Метрики кластерного анализа.
14. Методы определения близости между кластерами.
15. Иерархическая кластеризация. Дендограмма.
16. Метод k-средних.
17. Ассоциативные правила. Свойства антимонотонности.
18. Метрики построения ассоциативных правил.
19. Алгоритм построения ассоциативных правил a priori.
20. Общая характеристика деревьев решений.
21. Алгоритмы построения деревьев решений.
22. Оценка качества классификации. Задачи классификации. ROC-кривая. Таблица сопряженности.
23. Определение регрессионной модели. Логистическая регрессионная модель. Использование логистической модели для классификации.
24. Нейронные сети. Перцептрон. Радиальные базисные сети.
25. Использование карты Кохонена для решения задач классификации.
26. Общая характеристика QlikView.
27. Общая характеристика DeductorAcademic.
28. Общая характеристика средств интеллектуального анализа SQLServer.

Шкала оценивания

Обозначения		Формулировка требований к степени сформированности компетенции
Цифр.	Оценка	
2	Неуд.	Студент не имеет необходимых представлений о проверяемом материале
3	Удовл.	Знания не структурированы, на уровне ориентирования , общих представлений. Студент допускает неточности, приводит недостаточно правильные формулировки, нарушает логическую последовательность в изложении ответа на вопросы или в демонстрируемом действии.
4	Хор.	Знания, умения, навыки на аналитическом уровне. Компетенции в целом сформированные, но содержащие отдельные пробелы. Студент твердо знает материал, грамотно и по существу его излагает, однако допускает несущественные погрешности при ответе на заданный вопрос или в демонстрируемом действии.
5	Отл.	Знания, умения, навыки на системном уровне. Студент глубоко и прочно усвоил программный материал, исчерпывающе, последовательно и четко его излагает, умеет тесно

	увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, в том числе при видоизменении и решении нестандартных практических задач, правильно обосновывает принятое решение.
--	---

6. Методические материалы для освоения дисциплины

Экзамены организуются в период сессии в соответствии с текущим графиком учебного процесса. Продолжительность экзамена для каждого студента не может превышать четырех академических часов. Экзамен не может начинаться ранее 9.00 часов и заканчиваться позднее 21.00 часа. Экзамен проводится в аудитории, в которую запускаются одновременно не более 5 человек. Время на подготовку ответов по билету каждому обучающемуся отводится 30-40 минут. При явке на экзамен обучающийся должен иметь при себе зачетную книжку. Во время экзамена обучающиеся по решению преподавателя могут пользоваться учебной программой дисциплины и справочной литературой.

Методические рекомендации по работе над конспектом лекций во время и после проведения лекции.

Обучающимся рекомендуется в ходе лекционных занятий выполнять следующее: вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации по их применению, задавать преподавателю вопросы.

Целесообразно в конспектах лекций рабочих конспектах формировать поля, на которых возможно делать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных положений.

Методические рекомендации к семинарским (практическим) занятиям.

На семинарских занятиях в зависимости от темы занятия выполняется поиск информации по решению проблем, выработка индивидуальных или групповых решений, решение задач, итоговое обсуждение с обменом знаниями, участие в круглых столах, разбор конкретных ситуаций, командная работа, решение индивидуальных тестов, участие в деловых играх.

При подготовке к семинарским занятиям каждый обучающийся должен: изучить рекомендованную учебную литературу; подготовить ответы на все вопросы семинара.–

При подготовке к семинарским занятиям необходимо обратить внимание на виды работ, которые определены заданием. Существенный акцент делается на умение студента выполнять индивидуальные письменные задания, а также на работу студента с большим объемом информации, как в электронном, так и в печатном виде.

При подготовке к семинарским занятиям важно проработать материал лекций по конкретной теме, ознакомиться с указанной литературой и выполнить все необходимые

практические задания. Для семинарских занятий лучше завести отдельную папку с файлами или тетрадь со съемными листами для удобства работы.

Подготовка к контрольным мероприятиям.

При подготовке к контрольным мероприятиям обучающийся должен освоить теоретический материал, повторить материал лекционных и практических занятий, материал для самостоятельной работы по указанным преподавателям темам.

Самостоятельная работа осуществляется в виде изучения литературы, эмпирических данных по публикациям и конкретных ситуаций, подготовке индивидуальных работ, работа с лекционным материалом, самостоятельное изучение отдельных тем дисциплины; поиск и обзор учебной литературы, в т.ч. электронных источников; научной литературы, справочников и справочных изданий, нормативной литературы и информационных изданий.

7. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет"

7.1. Основная литература

1. Миркин Б.Г. Введение в анализ данных [Электронный ресурс]: учебник и практикум / Б. Г. Миркин; Нац. исслед. ун-т Высш. шк. экономики. - Электрон. дан. - М. : Юрайт, 2016. - 174 с.
2. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям : [хранилища данных и OLAP, очистка и предобработка данных, основные алгоритмы DataMining, сравнение и ансамбли моделей, решение бизнес задач на аналитической платформе Deductor] : учеб. пособие / Н. Паклин, В. Орешков. - 2-е изд., испр. - СПб.[и др.]: Питер, 2013. - 701

7.2. Дополнительная литература

1. Наумов В.Н. Средства бизнес-аналитики. – СПб.: СЗИУ, 2016.
2. Наумов В.Н. Проектирование информационных систем: Учебное пособие. - СПб.: СЗИУ, 2018.

7.3. Нормативные правовые документы и иная правовая информация

1. Правовая система «Гарант-Интернет» [Электронный ресурс]. – Режим доступа: <http://www.garweb.ru>.
2. Правовая система «КонсультантПлюс» [Электронный ресурс]. – Режим доступа: <http://www.consultant.ru>.

7.4. Интернет-ресурсы

1. Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
2. <http://www.kremlin.ru> - официальный сайт Президента Российской Федерации
3. <http://www.gov.ru> - сервер органов государственной власти Российской Федерации
4. <http://www.ach.gov.ru> - официальный сайт Счетной палаты Российской Федерации
5. <http://www.mvdinfor.ru> - официальный сайт Министерства внутренних дел Российской Федерации
6. <http://www.minfin.ru> - официальный сайт Министерства финансов Российской Федерации
7. <http://www.rosfinnadzor.ru> - официальный сайт Федеральной службы финансово-бюджетного надзора
8. <http://www.nalog.ru> - официальный сайт Федеральной налоговой службы
9. <http://www.castoms.ru> - официальный сайт Федеральной таможенной службы
10. <http://www.fedsfm.ru> - официальный сайт Федеральной службы по финансовому мониторингу.
11. <http://www.ffms.ru> - официальный сайт Федеральной службы по финансовым рынкам

7.5. Иные источники

Не используются

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

Курс включает использование программного обеспечения Microsoft Excel, Microsoft Word, Microsoft Power Point для подготовки текстового и табличного материала, графических иллюстраций.

Методы обучения предполагают использование информационных технологий (компьютерное тестирование, демонстрация мультимедийных материалов).

Задействованы Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).

Допускается применение системы дистанционного обучения.

№ п/п	Наименование
1.	Специализированные залы для проведения лекций:
2.	Специализированная мебель и оргсредства: аудитории и компьютерные классы, оборудованные посадочными местами
3.	Технические средства обучения: Персональные компьютеры; компьютерные проекторы; звуковые динамики; программные средства, обеспечивающие просмотр видеофайлов