

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Хоружий Людмила Ивановна
Должность: Директор института экономики и управления АПК
Дата подписания: 21.03.2024 11:48:48
Уникальный программный ключ:
1e90b132d9b04dce67585160b015dddf2cb1e6a9

 УТВЕРЖДАЮ:
Директор Института экономики и
управления АПК


Л.И. Хоружий
2023 г.

**Лист актуализации рабочей программы дисциплины
«Б1.В.ДВ.06.02 Интеллектуальный анализ данных и статистика»**

для подготовки бакалавров
Направление: 09.03.02 Информационные системы и технологии
Направленность: Компьютерные науки и интеллектуальный анализ данных
(Computer Science & Data Mining)
Форма обучения очная
Год начала подготовки: 2022
Курс 4
Семестры 7, 8

1. В рабочую программу не вносятся изменения. Программа актуализирована для 2023 г. начала подготовки.
2. Программа будет распространена при организации учебного процесса на направленность (профиль): Компьютерные науки и интеллектуальный анализ данных.

Разработчик (и): Демичев В.В., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)



(подпись)

«28» августа 2023 г.

Рабочая программа пересмотрена и одобрена на заседании кафедры статистики и кибернетики протокол № 11 от «28» августа 2023 г.

И.о. заведующего кафедрой статистики и кибернетики _____  А.В. Уколова

Лист актуализации принят на хранение:

И.о. заведующего кафедрой статистики и кибернетики _____  А.В. Уколова



МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –
МСХА имени К.А. ТИМИРЯЗЕВА»
(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)

Институт экономики и управления АПК
Кафедра статистики и кибернетики

УТВЕРЖДАЮ:
Директор института экономики и управ-
ления АПК
Л.И. Хоружий
“ 11 ” 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.ДВ.06.02 Интеллектуальный анализ данных и статистика

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»

Направленность:

Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)

Курс 4

Семестр 7,8

Форма обучения очная

Год начала подготовки 2022

Москва, 2022

Разработчик (и): Харитонов А.Е., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

«26» 08 2022г.

Рецензент: Коломеева Е.С., к.э.н.

(ФИО, ученая степень, ученое звание)

(подпись)

«26» 08 2022г.

Программа составлена в соответствии с требованиями ФГОС ВО, профессионального стандарта и учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

Программа обсуждена на заседании кафедры статистики и кибернетики протокол № 11 от «26» августа 2022 г.

И.о.зав. кафедрой Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

(подпись)

«26» 08 2022г.

Согласовано:

Председатель учебно-методической комиссии института экономики и управления АПК
Корольков А.Ф., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

(подпись)

«26» 08 2022г.

И.о.заведующего выпускающей кафедрой
статистики и кибернетики

Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

(подпись)

«26» 08 2022г.

/Заведующий отделом комплектования ЦНБ

(подпись)

СОДЕРЖАНИЕ

АННОТАЦИЯ	4
1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	5
2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ	5
3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	5
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	5
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ	6
ПО СЕМЕСТРАМ	6
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	11
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	13
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	17
6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	18
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ	18
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ.....	31
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	32
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	32
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	ОШИБКА! ЗАКЛАДКА НЕ ОПРЕДЕЛЕНА.
7.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ, РЕКОМЕНДАЦИИ И ДРУГИЕ МАТЕРИАЛЫ К ЗАНЯТИЯМ.....	32
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ	33
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ.....	33
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ	34
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ..	35
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ.....	35

Аннотация

рабочей программы учебной дисциплины

Б1.В.ДВ.06.02 «Интеллектуальный анализ данных и статистика»
для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»

Цель освоения дисциплины: Целью дисциплины «Интеллектуальный анализ данных и статистика» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в проведении анализ данных с использованием информационных технологий для осуществления деловой коммуникации в устной и письменной формах.

Место дисциплины в учебном плане: включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

Требования к результатам освоения дисциплины: в результате освоения дисциплины формируются следующие компетенции (индикаторы): УК-4(УК-4.2, УК-4.3); ПКос-8 (ПКос-8.1, ПКос-8.2, ПКос-8.3); ПКос-9 (ПКос-9.2, ПКос-9.3)

Краткое содержание дисциплины:

Предмет и метод статистики. Абсолютные, средние и относительные величины. Проверка статистических гипотез. Понятие, сущность и задачи интеллектуального анализа данных. Стадии интеллектуального анализа данных. Решение задач классификации и регрессии. Поиск ассоциативных правил в процессе анализа данных. Кластерный анализ. Анализ текстовой информации и аналитика. Визуализация данных.

Общая трудоемкость дисциплины составляет: 5 зачетных единиц (180 часов).

Промежуточный контроль: зачет в 7 семестре, экзамен в 8 семестре.

1. Цель освоения дисциплины

Целью дисциплины «Интеллектуальный анализ данных и статистика» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в проведении анализ данных с использованием информационных технологий для осуществления деловой коммуникации в устной и письменной формах.

2. Место дисциплины в учебном процессе

Дисциплина «Интеллектуальный анализ данных и статистика» включена в часть, формируемую участниками образовательных отношений (дисциплина по выбору) учебного плана. Дисциплина «Интеллектуальный анализ данных и статистика» реализуется в соответствии с требованиями ФГОС ВО, профессионального стандарта, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Интеллектуальный анализ данных и статистика» являются «Линейная алгебра», «Математический анализ», «Математическая статистика», «Теория вероятностей», «Алгоритмизация и программирование», «Моделирование временных рядов на иностранном языке», «Моделирование временных рядов», «Анализ экономических данных с использованием современных информационных технологий на иностранном языке», «Анализ экономических данных с использованием современных информационных технологий» «Расчетно-графическая статистика в среде R на иностранном языке», «Расчетно-графическая статистика в среде R».

Рабочая программа дисциплины «Интеллектуальный анализ данных и статистика» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

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

Образовательные результаты освоения дисциплины обучающимся, представлены в таблице 1.

4. Структура и содержание дисциплины

4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 5 зач.ед. (180 часов), их распределение по видам работ семестрам представлено в таблице 2.

Требования к результатам освоения учебной дисциплины

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
1.	УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)	УК-4.2 Уметь: применять на практике деловую коммуникацию в устной и письменной формах, методы и навыки делового общения на русском и иностранном языках		вести монологическую и диалогическую речь в рамках изученных тем с учетом правил речевого общения в информационно-технологической и статистической сфере	
2.			Владеть: навыками чтения и перевода текстов на иностранном языке в профессиональном общении; навыками деловых коммуникаций в устной и письменной форме на русском и иностранном языках; методикой составления суждения в межличностном деловом общении на русском и иностранном языках			навыками деловых коммуникаций в устной и письменной форме, относящимися к информационно-технологической и статистической деятельности
3.	ПКос-8	Способность проводить анализ данных с использованием информационных технологий в области сельского хозяйства, экономики, бухгалтерского учета, статистики, финансов и др.	ПКос-8.1 Знать: основы технологии производства продукции сельского хозяйства; теорию и методологию дисциплин экономического профиля (экономика, бухгалтерский учет, статистика, финансы и др.); информационные технологии анализа данных; источники информации для профессиональной деятельности	Современные программные средства интеллектуального анализа данных способы и		

				средства сбора статистической информации для профессиональной деятельности		
4.			ПКос-8.2 Уметь: собирать информацию для проведения анализа; устанавливать причинно-следственные связи между признаками; выбирать и применять, в том числе с использованием современных информационных технологий, методы анализа данных; делать выводы на основе проведенного анализа данных		применять информационные технологии и статистические методы для сбора, обработки и анализа данных, устанавливать причинно-следственные связи между признаками с использованием статистических методов и интеллектуального анализа данных; интерпретировать полученные результаты	
5.			ПКос-8.3 Владеть: методологией и навыками проведения анализа данных с использованием информационных технологий в области экономики, в том числе экономики сельского хозяйства			навыками разработки информационных систем обработки данных в области экономики, в том числе экономики сельского хозяйства
6.	ПКос-9	Способен осуществлять разработку, отладку и рефакторинг	ПКос-9.2 Уметь: использовать методы и приемы формализации и алгоритмизации поставленных задач; использо-		применять языки программирования R и Python для	

		кода программного обеспечения, интеграцию программных модулей и компонент, в том числе взаимодействующих с внешней средой, средствами выбранных языков программирования	<p>вать программные продукты для графического отображения алгоритмов; применять стандартные алгоритмы в соответствующих областях; применять выбранные языки программирования для написания программного кода; использовать выбранную среду программирования; применять инструментарий для создания и актуализации исходных текстов программ; выявлять ошибки в программном коде, интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; применять методы и приемы отладки программного кода</p>		написания программного кода для проведения интеллектуального анализа данных	
7.			<p>ПКос-9.3 Владеть навыками; составления формализованных описаний решений поставленных задач в соответствии с требованиями технического задания; разработки алгоритмов решения поставленных задач в соответствии с требованиями технического задания или других принятых в организации нормативных документов; создания программного кода в соответствии с техническим заданием (готовыми спецификациями); оптимизации программного кода с использованием специализированных программных средств; анализа и проверки исходного программного кода; отладки программного кода на уровне программных модулей и межмодульных</p>			<p>навыками создания программного кода в соответствии с техническим заданием</p>

			взаимодействий и взаимодействий с окружением			
--	--	--	---	--	--	--

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость всего/*		
	час.	в т.ч. по семестрам	
		№ 7	№ 8
Общая трудоёмкость дисциплины по учебному плану	180,00/8	72,00/4	108,00/4
1. Контактная работа:	100,65/8	50,25/4	50,40/4
Аудиторная работа	100,65/8	50,25/4	50,40/4
<i>в том числе:</i>			
<i>лекции (Л)</i>	40,00	16,00	24,00
<i>практические занятия (ПЗ)</i>	58,00/8	34,00/4	24,00/4
<i>консультации перед экзаменом</i>	2,00	-	2,00
<i>контактная работа на промежуточном контроле (КРА)</i>	0,65	0,25	0,40
2. Самостоятельная работа (СРС)	79,35	21,75	57,60
<i>контрольная работа</i>	15,00	5,00	10,00
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям и т.д.)</i>	30,75	7,75	23,00
<i>Подготовка к зачету</i>	9,00	9,00	-
<i>Подготовка к экзамену</i>	24,60	-	24,60
Вид промежуточного контроля:		зачёт	экзамен

* в том числе практическая подготовка.

4.2 Содержание дисциплины

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнено)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	всего/*	ПКР	
Раздел 1. Статистика	25,75	4	16		5,75
Тема 1. Статистика и анализ данных	3,75	1	2		0,75
Тема 2. Абсолютные, средние и относительные величины	4	1	2		1
Тема 3. Проверка статистических гипотез.	12	2	8		2
Тема 4. Корреляционно-регрессионный анализ	6	-	4		2
Раздел 2. Интеллектуальный анализ данных	46	12	18		16
Тема 5. Понятие, сущность и задачи интеллектуального анализа данных	10	2	2		6
Тема 6. Решение задач классификации и регрессии	36/4	10	16/4		10
Контактная работа на промежуточном контроле (КРА)	0,25			0,25	
Всего за 7 семестр	72/4	16	34	0,25	21,75

Тема 7. Поиск ассоциативных правил в процессе анализа данных	22	6	4		12
Тема 8. Кластерный анализ	18/4	6	6/4		6
Тема 9. Нейросетевые методы анализа	26,6	4	6		16,6
Тема 10. Анализ текстовой информации и аналитика	20	4	4		12
Тема 11. Визуализация данных	19	4	4		11
Консультации перед экзаменом	2			2	
Контактная работа на промежуточном контроле (КРА)	0,4			0,4	
Всего за 8 семестр	108/4	24	24/4	2,4	57,6
Итого по дисциплине	180/8	40	58/8	2,65	79,35

* в том числе практическая подготовка

Раздел 1 Статистика

Тема 1. Статистика и анализ данных

Понятие о статистике как науке. Предмет статистики. Метод статистики. Статистическая закономерность. Основные категории и понятия статистики: статистическая совокупность, единица совокупности, признак, вариация, статистический показатель, система показателей. Различия между интеллектуальным анализом данных и статистикой.

Тема 2. Абсолютные, средние и относительные величины

Значение абсолютных и относительных величин для статистического анализа данных. Виды абсолютных величин, единицы измерения и способы получения. Средняя величина в статистике, ее сущность и условия применения. Относительные величины, их виды и способы расчета в пакетах статистических программ. Взаимосвязь абсолютных и относительных величин.

Тема 3. Проверка статистических гипотез.

Понятие выборочного наблюдения. Теоретические основы выборочного наблюдения. Генеральная и выборочная совокупности, их обобщающие характеристики. Средняя и предельная ошибки выборки. Понятие статистической гипотезы. Основные критерии для проверки статистических гипотез. Этапы проверки статистических гипотез. Критерий t-Стьюдента и F-Фишера. Тест ранговой корреляции Уилкоксона. Тест Шапиро-Уилкса. Тест Колмогорова-Смирнова. Дисперсионный анализ.

Тема 4. Корреляционно-регрессионный анализ

Генеральное уравнение парной и множественной линейной регрессии и их выборочные оценки. Интерпретация параметров уравнения регрессии. Показатели качества модели. Оценка достоверности параметров и коэффициента корреляции с использованием критерия Стьюдента. Оценка достоверности уравнения регрессии в целом, дисперсионный анализ, критерий Фишера.

Прогнозирование на основе уравнения регрессии. Построение нелинейных регрессий в среде R, Python. Оценка достоверности параметров нелинейных мо-

делей. Отбор факторов в модель множественной регрессии. Версия всех возможных регрессий и пошаговые процедуры отбора в среде R.

Раздел 2. Интеллектуальный анализ данных

Тема 5. Понятие, сущность и задачи интеллектуального анализа данных

История возникновения, основные понятия и главные задачи интеллектуального анализа данных. Взгляд на технологию Data Mining как на часть рынка информационных технологий. Стадии Data Mining: постановка цели и задач, сбор данных, создание базы данных, исследование и подготовка данных, разделение совокупности на сегменты, построение и проверка достоверности предсказательной модели, инсталляция модели, обучение пользователей, использующих модель, мониторинг, расширение модели.

Тема 6. Решение задач классификации и регрессии

Правила классификации и математические функции и методы их построения. Построение основных методов классификации в среде R и Python. Дерево решений. Метод опорных. Случайный лес. Логистическая регрессия. Дискриминантный анализ. Байесовская (наивная) классификация. Метод ближайшего соседа.

Тема 7. Поиск ассоциативных правил в процессе анализа данных

Понятия и методы выявления закономерностей в интеллектуальном анализе данных. Использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование). Анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях.

Тема 8. Кластерный анализ

Иерархические и неиерархические методы в кластерном анализе. Рассмотрение примеров использования кластерного анализа.

Тема 9. Нейросетевые методы анализа

Сущность и применение нейросетевых методов анализа. Классификаторы на основе искусственных нейронных сетей. Настройка параметров нейронных сетей средствами пакета caret.

Тема 10. Анализ текстовой информации и аналитика

Основные методы добычи и анализа текстовых данных с целью обнаружения закономерностей с акцентом на статистические подходы.

Тема 11. Визуализация данных

Методы и средства визуального представления информации, в частности, способы представления информации в одно-, двух-, трехмерном измерениях, а также способы отображения информации в более чем трех измерениях. Описание принципов качественной визуализации. Основные тенденции в области визуализации.

4.3 Лекции/ практические занятия

Таблица 4

Содержание практических занятий и контрольные мероприятия

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
1.	Раздел 1. Статистика				
	Тема 1. Статистика и анализ данных	Лекция № 1. Программные средства статистической обработки данных	ПКос-8.1	-	1
		Практическая работа № 1 Предмет, метод и задачи статистики	УК-4.2 УК -4.3 ПКос-8.1	Участие в обсуждении	2
	Тема 2. Абсолютные, средние и относительные величины	Лекция № 2. Возможности среды R для обработки данных	ПКос-8.1	-	1
		Практическое занятие №2 «Расчет основных показателей центральной тенденции и вариации в пакете R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
	Тема 3. Проверка статистических гипотез.	Лекция № 3. Статистические тесты, реализованные в пакетах R	ПКос-8.1	-	2
		Практическое занятие № 3. Статистические тесты в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	4
		Практическое занятие № 4. Разработка статистических тестов на языке R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	4
	Тема 4. Корреляционно-регрессионный анализ	Практическое занятие № 5. Корреляционно-регрессионный анализ в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Контрольная работа по темам 1, 2, 3 и 4.	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Письменная работа	2
2	Раздел 2 Интеллектуальный анализ данных				
	Тема 5. Понятие, сущность и задачи интеллектуального анализа данных	Лекция № 4. Понятие, сущность и задачи интеллектуального анализа данных	ПКос-8.1	-	2
		Практическое занятие №6 «Интеллектуальный анализ данных и статистика: в чём различия?»	УК-4.2 УК -4.3	Участие в обсуждении	2
	Тема 6. Решение задач классификации и регрессии	Лекция № 5. Классификация методов Data mining	ПКос-8.1		2
		Лекция № 6. Методы классификации в Data mining	ПКос-8.1		6
		Лекция № 7. Регрессия в Data mining	ПКос-8.1		2

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Практическое занятие №7 Построение деревьев решений в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №8 Метод опорных векторов в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №9 Случайный лес в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №10 Логистическая регрессия в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №11 Дискриминантный анализ в среде R	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №12 Байесовская (наивная) классификация;	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №13 Метод ближайшего соседа;	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Контрольная работа по темам 5 и 6	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Контрольная работа (письменная работа)	2
	Тема 7. Поиск ассоциативных правил в процессе анализа данных	Лекция 8. Поиск ассоциативных правил в процессе анализа	ПКос-8.1	-	6
		Практическое занятие №14 «Методы поиска ассоциативных правил»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
		Практическое занятие №15 «Алгоритм Apriori»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
	Тема8. Кластерный анализ	Лекция 8. Кластерный анализ в среде R	ПКос-8.1	-	6

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Практическое занятие №16 «Иерархические методы в кластерном анализе»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2/2
		Практическое занятие №17 «Не-иерархические методы в кластерном анализе»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	4/2
	Тема 9. Нейросетевые методы анализа	Лекция 9. Нейросетевые методы анализа	ПКос-8.1	-	4
		Практическое занятие №18 Классификаторы на основе искусственных нейронных сетей	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	3
		Практическое занятие №19 Настройка параметров нейронных сетей средствами пакета caret	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	3
	Тема 10. Анализ текстовой информации и аналитика	Лекция 8. Анализ текстовой информации и аналитика	ПКос-8.1	-	4
		Практическое занятие №20 «Анализ текстовой информации: подходы, особенности, перспективы»	УК-4.2 УК -4.3	Участие в обсуждении	2
		Практическое занятие №21 «Типичные задачи анализа текстовой информации: категоризация, кластеризация текстов, смысловой анализ»	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Защита работы	2
	Тема 11. Визуализация данных	Лекция 9. Методы и средства визуализации данных	ПКос-8.1		4
		Практическое занятие №22 «Классификация методов визуализации данных»	УК-4.2 УК -4.3 ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Участие в обсуждении Защита работы	2
		Практическое занятие №23 «Программное обеспечение для визуализации данных»	УК-4.2 УК -4.3	Участие в обсуждении	1
		Итоговая контрольная работа	ПКос-8.2 ПКос-8.3 ПКос-9.2 ПКос-9.3	Тестирование	1

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
Раздел 1. Статистика		
1.	Тема 1. Статистика и анализ данных	Закон больших чисел и его значение для статистики Статистическое наблюдение и его организация в зарубежных странах (УК -4.3, УК -4.3, ПКос-8.1)
2.	Тема 2. Абсолютные, средние и относительные величины	Условия применения относительных величин Среднее арифметическое и его свойства Мода, медиана, квартили (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
3	Тема 3. Проверка статистических гипотез.	Статистические гипотезы, используемые в реальных примерах. Типы отбора проб. Расчет погрешности выборки. Ограничение t-теста (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
4	Тема 4. Корреляционно-регрессионный анализ	Корреляционный и регрессионный анализ в пакетах прикладных статистических программ (ПКос-8.1, ПКос-8.2, ПКос-8.3, ПКос-9.2, ПКос-9.3))
Раздел 2. Интеллектуальный анализ данных		
5	Тема 5. Понятие, сущность и задачи интеллектуального анализа данных	Области практического применения интеллектуального анализа данных. Методология сбора данных Основные типы баз данных (УК -4.3, УК -4.3, ПКос-8.1)
6	Тема 6. Решение задач классификации и регрессии	Одно- и многомерная классификация Валидация моделей прогнозной регрессии Использование регрессионного анализа в различных прикладных областях (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
7	Тема 7. Поиск ассоциативных правил в процессе анализа данных	AIS –алгоритм. SETM-алгоритм. DIC-алгоритм (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
8	Тема 8. Кластерный анализ	Программное обеспечение для кластеризации и сегментации Меры расстояний в кластерном анализе Анализ последовательности (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
9	Тема 9. Нейросетевые методы анализа	Сферы применения нейросетевых методов анализа (ПКос-8.1, ПКос-8.2, ПКос-9.2, ПКос-9.3))
10	Тема 10. Анализ текстовой информации и аналитика	Текст Mining в R. Статистические методы для добычи текста Методы добычи контекстного текста (УК -4.3, УК -4.3, ПКос-8.1)
11	Тема 11. Визуализация данных	Современные тенденции в визуализации данных Примеры интерактивной визуализации данных Примеры визуализации статических данных 3D визуализация (УК -4.3, УК -4.3, ПКос-8.1, ПКос-9.2, ПКос-9.3))

5. Образовательные технологии

Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий
1.	Классификация методов Data mining	Л
2.	Практическое занятие № 4. Разработка статистических тестов на языке R	ПЗ

6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины

6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

Контрольная работа №1

Вариант 1

1. Создать вектор v (2,5,8,2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2) Создать вектор b из 20 случайных чисел в диапазоне от 1 до 100 без повторов.

Найти: для вектора b : среднюю (по формуле и по встроенной функции), среднеквадратическое отклонение (округлить в большую сторону), минимальное и максимальное значение, количество элементов.

Вектор v отсортировать в порядке возрастания. Найти значения элементов по ранжированному ряду с номерами 2 и 8.

Рассчитать ковариацию между векторами v и b ($cov = \overline{xy} - \bar{x} * \bar{y}$).

2. Проверить по данным базы iris равенство средних по переменной Sepal.Length для сортов setosa и versicolor. $c = t$ фактическое Сделать вывод.

3. Провести тест на независимость выборок для данных:

	A	B
C	30	15
D	10	13

4. По данным sleep построить регрессию зависимости лекарства (group) от продолжительности сна (extra). Сделать вывод.

5. По данным mtcars построить зависимость топливной экономичности (mpg) от веса(wt) и вместимостью двигателя (disp). Сделать вывод.

Вариант 2

1. Создать вектор v (2,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8,5,8)

Создать вектор b из 30 случайных чисел в диапазоне от -100 до 100 без повторов.

Найти: для вектора b : медиану (округлить до целого числа), дисперсию (округлить в большую сторону), сумму элементов, произведение элементов, количество элементов.

Вектор v отсортировать в порядке возрастания. Найти значения элементов по ранжированному ряду с номерами 4 и 15.

Рассчитать коэффициент корреляции между векторами v и b ($r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$).

2. Сравнить на соответствие нормальному закону по критерию Шапиро-Уилка базы iris переменной Sepal.Length. $h = W$

3. Проверить равенство средних для групп:

№ группы	x_i
1	2
1	8
1	6
1	5
1	7
2	4
2	3
2	5
2	6
2	5
3	3
3	4
3	3
3	6
3	5

4. Рассчитать несмещенную оценку дисперсии для №+10 переменных Sepal.Length сорта versicolor. № - номер Вашей команды. $i = S^2$

5. По данным sleep построить регрессию зависимости лекарства (group) от прироста продолжительности сна (extra). Сделать вывод.

Контрольная работа № 2

Вариант 1

По данным о студентах предсказать поступит он или нет методом «случайный лес» и методом опорных векторов.

Вариант 2 .

По данным о студентах предсказать поступит он или нет методом опорных векторов с ядерной функцией и методом «дерево решений»

Вариант 3

По данным о студентах предсказать поступит он или нет с использованием нейронных сетей и с помощью Байесовской классификации.

Пример тестовых заданий

1. Что такое интеллектуальный анализ данных?

- a) извлечение информации из необработанных данных
- b) хранение и обработка информации в базах данных
- c) визуализация данных
- d) статистический метод обработки данных

2. Какая из концепций является базисной в технологии интеллектуального анализа данных?

- a) концепция естествознания
- b) концепция управления
- c) концепция закономерностей
- d) концепция формирования

3. Что является результатом представления знаний в интеллектуальном анализе данных?

- a) закономерность
- b) проект
- c) модель
- d) тренд

4. Задача классификации – это _____ задача.

- a) описательная
- b) предсказательная
- c) качественная
- d) количественная

5. Задача кластеризации – это _____ задача.

- a) описательная
- b) предсказательная
- c) качественная
- d) количественная

6. Единицы наблюдения, которые значительно отличаются от большинства других единиц в наборе данных:

- a) транзакция
- b) порядковое число
- c) интервалы
- d) резко выделяющиеся значения

7. Набор конкретных примеров с известным исходом:

- a) клиент-сервер
- b) классификатор
- c) учебный набор
- d) интеллектуальный анализ данных

8. Преобразование данных включает в себя...

- a) разделение данных из одного источника на несколько источников данных
- b) процесс изменения данных с обобщенного уровня на более детализированный
- c) объединение данных из одного источника с другими источниками данных
- d) процесс изменения данных с детализированного уровня на более обобщенный

9. _____ - полезный метод обнаружения закономерностей в начале процесса интеллектуального анализа данных.

- a) расчёт меры расстояния
- b) дерево принятия решений
- c) ассоциативные правила
- d) приёмы визуализации

10. _____ - класс моделей, принцип которых основан на аналогии с работой человеческого мозга:

- a) нейронные сети
- b) кластеры
- c) дерево принятия решений
- d) правило классификации

11. Какая иерархическая структура у деревьев принятия решений?

- a) ЕСЛИ... ТО...
- b) НИ... НИ...
- c) ЛИБО... ЛИБО...
- d) КАК... ТАК И...

12. Что из перечисленного является математическим уравнением, связывающим переменные x и y ?

- a) регрессия
- b) интерполяция
- c) кластеризация
- d) экстраполяция

13. Уравнение вида $y = a + bx$:

- a) полиномиальное уравнение
- b) линейная регрессия
- c) регрессия
- d) интерполяция

14. Какой из этих показателей является измерением качества модели регрессии?

- a) средняя арифметическая
- b) дисперсия

- c) стандартное отклонение
- d) коэффициент детерминации

15. Коэффициент корреляции всегда лежит между значениями...

- a) 0 и 1
- b) -1 и 1
- c) -1 и 0
- d) 0 и 2

16. Изучение взаимосвязи между несколькими переменными – это задача...

- a) парной регрессии
- b) множественной регрессии
- c) дерева принятия решения
- d) моделирования

17. Что из перечисленного является методом построения правил классификации?

- a) 1R-алгоритм
- b) метод Naive Bayes
- c) оба варианта верны
- d) ни один из вариантов не является верным

18. Ожидаемое значение y , когда $X = 0$, равно:

- a) коэффициенту полной регрессии
- b) условному началу
- c) коэффициенту корреляции
- d) коэффициенту детерминации

19. Собранные в разные моменты времени значения каких-либо параметров - это...

- a) панельные данные
- b) пространственные данные
- c) данные временного ряда
- d) ни один из вариантов

20. Изучение взаимосвязи между двумя переменными возможно благодаря...

- a) двумерному графику
- b) гистограммам
- c) графикам временных рядов
- d) ни один из вариантов

21. Главная цель поиска ассоциативных правил заключается в том, чтобы...

- a) создать правила классификации

- b) проверить достоверность регрессионной модели
- c) определить главную идею той или иной текстовой информации
- d) выявить закономерности между связанными событиями в базах данных

22. Если несколько событий связаны друг с другом, то это...

- a) ассоциация
- b) последовательность
- c) классификация
- d) кластеризация

23. Отношение транзакций, которые имеют набор F (D_F) к общему количеству транзакций (D) называется...

- a) весомым уровнем набора F
- b) моделью набора F
- c) кластером набора F
- d) уровнем поддержки набора F

24. Набор предметов (F) называется частым, когда...

- a) $\text{Supp}(F) < \text{Supp}(\min)$
- b) $\text{Supp}(F) > \text{Supp}(\min)$
- c) $\text{Supp}(F) = \text{Supp}(\min)$
- d) ни один из вариантов

25. Объекты некоторого набора предметов, которые подвергаются анализу, называются...

- a) пропорции
- b) центроиды
- c) транзакции
- d) модели

26. Что является последовательностью в поиске ассоциативных правил?

- a) конечные действительные числа
- b) ранжированный ряд
- c) упорядоченное множество некоторых объектов
- d) ни один из вариантов

27. Что из перечисленного не является оценкой полезности ассоциативных правил?

- a) качество
- b) поддержка
- c) достоверность
- d) улучшение

28. Поддержка любого набора объектов не может превышать минимальной поддержки любого из его подмножеств. Это основное свойство...

- a) метода Naive Bayes
- b) алгоритма Apriori
- c) дерева принятия решений
- d) кластеризации

29. Один из методов кластерного анализа называется...

- a) стандартное отклонение
- b) k-средние
- c) регрессия
- d) дисперсия

30. Какой из алгоритмов является иерархическим?

- a) агломеративный
- b) метод k-средних
- c) метод нечеткой кластеризации C-средних
- d) ни один из вариантов

31. Что из перечисленного не является методом пересчёта расстояний между кластерами?

- a) расстояние между ближайшими соседями
- b) расстояние между дальними соседями
- c) метод медиан
- d) модальный метод

32. Какой из алгоритмов является неиерархическим?

- a) агломеративный
- b) дивизимный
- c) метод k-средних
- d) ни один из вариантов

33. Что такое кластеры?

- a) данные временных рядов
- b) однородные группы объектов
- c) разнородные группы объектов
- d) статистическая совокупность

34. Одной из мер близости, используемой в кластеризации, является...

- a) расстояние Чебышева
- b) Евклидово расстояние
- c) оба варианта верны
- d) ни один из вариантов не является верным

35. Общепринятый способ визуализации результатов кластерного анализа является построение...

- a) гистограммы

- b) двумерного графика
- c) дендрограммы
- d) графика временного ряда

36. Что такое качество кластеризации?

- a) степень приближения результата кластеризации к идеальному решению
- b) правильный выбор количества кластеров
- c) мера идеального расстояния между кластерами
- d) пригодность полученных результатов для дальнейшего исследования

37. Что из перечисленного не является подходом к оценке поисковых информационных систем?

- a) полнота (recall)
- b) выпадение (fall-out)
- c) закономерности (patterns)
- d) точность (precision)

38. Что является первым этапом в анализе текстовой информации?

- a) предварительная обработка документов
- b) извлечение информации из текста
- c) интерпретация результатов
- d) поиск информации

39. Что такое стемминг?

- a) морфологический поиск
- b) поисковая система
- c) текстовый документ
- d) интересная закономерность

40. Классификация документов является синонимом к слову...

- a) стемминг
- b) аннотирование
- c) категоризация
- d) ни один из вариантов

41. Одной из задач анализа текстовой информации является...

- a) кластеризация
- b) извлечение ключевых понятий
- c) классификация
- d) всё вышеперечисленное

42. Что является примером текстовых данных?

- a) веб-страницы
- b) e-mail
- c) нормативные документы
- d) всё вышеперечисленное

43. Слова, которые являются вспомогательными и несут мало информации о содержании документа, - это...

- a) N-граммы
- b) стоп-слова
- c) диалекты
- d) слова из веб-страниц

44. Отношением числа релевантных документов, найденных информационно-поисковой системой, к общему числу найденных документов называется...

- a) точностью
- b) полнотой
- c) условием
- d) эффектом

45. График может быть удобным представлением данных, если...

- a) существует взаимосвязь между объектами данных
- b) объекты данных показывают определенную тенденцию
- c) оба варианта верны
- d) ни один из вариантов не является верным

46. Первым этапом визуализации данных является...

- a) беглый обзор
- b) построение графика
- c) интерпретация результатов
- d) оценка эффективности

47. Одним из методов визуализации является...

- a) кластеризация
- b) категоризация
- c) геометрические преобразования
- d) ни один из вариантов

48. Что из перечисленного не является графиком?

- a) гистограмма
- b) круговая диаграмма
- c) ранжированный ряд
- d) лепестковая диаграмма

49. Визуализация данных позволяет нам обнаружить...

- a) закономерности
- b) тренды
- c) корреляции
- d) всё вышеперечисленное

Примерные темы для обсуждения

Практическая работа № 1 Предмет и метод и задачи статистики

1. Определение понятия «статистика».
2. Чем вызвано появление и развитие науки и практики статистики?
3. Какие явления изучает статистика, что является объектом ее исследования?
4. Охарактеризуйте понятие «признак» и опишите его виды.
5. Чем различаются понятия «показатель» и «статистический показатель»?
6. Укажите отличительные особенности статистических показателей.
7. Что является предметом статистики?
8. Раскройте содержание системы статистических показателей.
9. В чем состоит сущность статистических закономерностей?
10. Охарактеризуйте основные черты метода статистики и этапы статистического исследования.
11. Как организована статистика в мире и в современной России?

Практическое занятие №6 «Интеллектуальный анализ данных и статистика: в чём различия?»

1. Понятие интеллектуального анализа данных.
2. Основные этапы развития интеллектуального анализа данных.
3. Основные задачи интеллектуального анализа данных.
4. Описательные и предсказательные задачи.
5. Какие статистические методы могут применяться в Data Mining.
6. Основные области практического применения интеллектуального анализа данных.
7. Как интеллектуальный анализ данных используется экономистами?

Практическое занятие №20 «Анализ текстовой информации: подходы, особенности, перспективы»

1. Основные этапы текстового анализа.
2. Задачи текстового анализа.
3. Извлечение ключевых понятий из текста.
4. Классификация документов.
5. Кластерный анализ документов.
6. Существующие программные обеспечения в области анализа текстовой информации.

Практическое занятие №22 «Классификация методов визуализации данных»

1. Характеристики инструментов визуализации данных.
2. Методы визуализации.
3. Существующие тренды в области визуализации данных.
4. 3D визуализация.

Практическое занятие №23 «Программное обеспечение для визуализации данных»

1. OfficeReports
2. Second Prism
3. Databoard
4. DataMarket
5. Q Research Software
6. Statwing

Пример работ

Практическое занятие №2 «Расчет основных показателей центральной тенденции и вариации в пакете R

Загрузить исходные данные в RStudio. Написать функцию, для расчета среднего значения, моды и медианы, дисперсии, размаха вариации и коэффициента вариации для ранжированного, дискретного и интервального рядов распределения. Рассчитать эти показатели с использованием встроенных функций. Сравнить значения. Оформить отчет с выводами.

Практическое занятие № 3. Статистические тесты в среде R

Загрузить исходные данные в RStudio.

1. Проверить соответствие исходных данных нормальному закону распределения с использованием критериев Колмогорова-Смирнова, Уилкоксона, Шапиро-Уилка, Пирсона.
 2. Сравнить различия в средних по критерию Стьюдента, F-Фишера.
 3. Построить таблицу частот по двум признакам. Проверить независимость выборок по критерию χ^2 – Пирсона.
- Оформить отчет с выводами.

Практическое занятие № 4. Разработка статистических тестов на языке R

По исходным данным предыдущей задачи написать функцию для применения дисперсионного анализа, критерия t-Стьюдента. Для дисперсионного анализа предусмотреть процедуру попарного сравнения методом Q-Тьюки. Оформить отчет с выводами.

Практическое занятие № 5. Корреляционно-регрессионный анализ в среде R

Загрузить исходные данные в RStudio. Построить уравнение парной линейной регрессии. Сделать прогноз по уравнению. Построить множественное уравнение линейной регрессии. Сделать прогноз по уравнению. Вывести summary. Оформить отчет с выводами.

Вопросы к зачету

1. Общее определение статистики
2. Предмет статистики
3. Метод статистики
4. Основные понятия статистики

5. Различия между сбором данных и статистикой
6. Типы единиц измерения, используемые для измерения абсолютных статистических показателей.
7. Типы относительных величин. Процедура их расчета.
8. Что такое среднее значение? Экспоненциальные средства и порядок их расчета.
9. Структурные средства и порядок их расчета.
10. Что такое система статистических показателей?
11. Основные виды статистических проверок гипотез.
12. Критерий уровня значимости.
13. Общий метод построения доверительных интервалов и проверки статистических гипотез.
14. Смысл понятий «ошибка первого рода», «ошибка второго рода», «сила теста».
15. Критические ограничения для единичных и двойных критериев для данного уровня значимости.
16. Приведите примеры практических задач при проверке гипотез о равенстве ожидания, дисперсии.

Вопросы к экзамену

1. Определение интеллектуального анализа данных.
2. Основные этапы разработки интеллектуального анализа данных.
3. Основные задачи интеллектуального анализа данных.
4. Описательные и прогностические задачи.
5. Регресс и классификация задач.
6. Ассоциативные правила.
7. Задача кластеризации.
8. Основные области практического применения интеллектуального анализа данных.
9. Как интеллектуальный анализ данных используется экономистами.
10. Определение цели проекта интеллектуального анализа данных.
11. Сбор данных для реализации проекта интеллектуального анализа данных.
12. Создание базы данных для хранения и обработки данных.
13. Изучение и подготовка данных.
14. Составление и проверка прогностических моделей.
15. Развертывание моделей.
16. Обучение пользователей моделям.
17. Мониторинг и обогащение моделей.
18. Способы представления отношений между переменными.
19. Правила классификации.
20. Деревья решений.
21. Математические функции.
22. 1R-алгоритм.
23. Наивный байесовский метод.
24. ID3 и C4.5 алгоритмы для построения деревьев решений.

25. Линейные методы построения уравнений описания отношений между переменными.
26. Метод наименьших квадратов (МНК).
27. Нелинейные методы.
28. Постановка задачи правил ассоциации.
29. Анализ последовательности.
30. Представление результатов Data Mining правил ассоциации.
31. Априорный алгоритм.
32. Постановка проблемы кластеризации.
33. Меры расстояний в кластерном анализе: евклидово расстояние, расстояние Чеби-Шева.
34. Представление результатов кластерного анализа.
35. Классификация алгоритмов кластеризации.
36. Иерархические алгоритмы: агломерационные и дивизионные методы кластеризации.
37. Неиерархические алгоритмы: k-means, Fuzzy C-Means.
38. Основные этапы анализа текста.
39. Задачи добычи текста.
40. Извлечение центральных понятий из текста.
41. Классификация текстовых документов.
42. Кластеризация текстовых документов.
43. Пакеты программ для анализа текста.
44. Характеристика средств визуализации данных.
45. Методы визуализации.
46. Основные тенденции в визуализации данных.
47. 3D визуализация.

6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания

Текущий контроль знаний, умений и навыков проводится в форме тестирования и контрольных работ с расчетными задачами и теоретическими вопросами, а также предусмотрена защита индивидуальных задач, выполняемых каждым студентом на практических занятиях. Ликвидация студентами текущих задолженностей производится также в форме выполнения индивидуальной задачи по соответствующей теме и дальнейшей ее защиты преподавателю кафедры.

Итоговая оценка учитывает результаты рейтинговой системы контроля знаний (вклад 50%), результаты контрольных работ по всем темам (вклад 50%) в VII семестре, результаты ответа на экзаменационные вопросы в VIII семестре. Вид промежуточного контроля по данному направлению: VII семестр – зачет; VIII семестр – экзамен. Критерии выставления оценок по системе:

0-59 % от максимального количества баллов – «незачтено»;

60 и более %– «зачтено».

Критерии выставления оценок на экзамене по четырехбалльной системе:

Неудовлетворительно – 0-59% от максимального количества баллов;
Удовлетворительно – 60 – 74%;
Хорошо – 75 – 84%;
Отлично – 85 – 100%.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1 Основная литература

1. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206711> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
2. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 228 с. — ISBN 978-5-8114-8519-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176662> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
3. Митина, О. А. Языки программирования для статистической обработки данных (R) : учебное пособие / О. А. Митина. — Москва : РТУ МИРЭА, 2020. — 191 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/163912> (дата обращения: 15.11.2022). — Режим доступа: для авториз. пользователей.
4. Алексеев, Д. С. Технологии интеллектуального анализа данных : учебник для вузов / Д. С. Алексеев, О. В. Щекочихин. — Санкт-Петербург : Лань, 2022. — 176 с. — ISBN 978-5-8114-8299-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/187559> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.
5. Демидова, Л. А. Интеллектуальный анализ данных на языке Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2021. — 92 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/218693> (дата обращения: 14.12.2022). — Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Перекаатов, А. С. Статистическая обработка экспериментальных данных. Полный факторный эксперимент в языке R : учебное пособие / А. С. Перекаатов, М. Б. Никифоров. — Рязань : РГРТУ, 2019. — 48 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/168309> (дата обращения: 28.11.2022). — Режим доступа: для авториз. пользователей.
2. Гришин, В. А. Методы обработки данных и моделирование на языке R : учебно-методическое пособие / В. А. Гришин, М. С. Тихов. — Нижний Новго-

род : ННГУ им. Н. И. Лобачевского, 2019. — 54 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/144653> (дата обращения: 15.11.2022). — Режим доступа: для авториз. пользователей.

3.

4. Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/118287> (дата обращения: 15.11.2022). — Режим доступа: для авториз. Пользователей.

5. Гильванов, Р. Г. Технологии обработки информации: Лабораторный практикум : учебное пособие / Р. Г. Гильванов, , А. В. Забродин. — Санкт-Петербург : ПГУПС, 2022. — 45 с. — ISBN 978-5-7641-1819-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/279065> (дата обращения: 28.11.2022). — Режим доступа: для авториз. пользователей.

6. Зыков, С. В. Программирование : учебник и практикум для вузов / С. В. Зыков. — Москва : Издательство Юрайт, 2022. — 320 с. — (Высшее образование). — ISBN 978-5-534-02444-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489754>

7.3 Методические указания, рекомендации и другие материалы к занятиям

1. Харитонов, А.Е. Хранилища и системы интеллектуального анализа данных: методические указания / А.Е. Харитонов. – М.: РГАУ-МСХА им. К.А. Тимирязева, 2016. – 25 с.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)
2. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)
3. Анаконда. URL: <https://www.anaconda.com/distribution/> (открытый доступ)
4. Официальный сайт Росстата. URL: <http://www.gks.ru/> (открытый доступ)
5. Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru> (открытый доступ)
6. Bureau of Economic Analysis. URL: <http://www.bea.gov> (открытый доступ)
7. Московская международная валютная биржа. <http://www.micex.ru> (открытый доступ)
8. Официальный сайт Всемирного банка . URL: [http:// www.worldbank.org](http://www.worldbank.org) (открытый доступ)

9. Официальный сайт Министерства финансов РФ. URL: <http://www.minfin.gov.ru> (открытый доступ)

10. Официальный сайт Национального бюро экономических исследований США. URL: <http://www.nber.org> (открытый доступ)

9. Перечень программного обеспечения и информационных справочных систем

Таблица 9

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	1-11	R	расчётная	r-project	2022
2	1-11	RStudio	расчётная	r-project	2022
3	1-11	Anaconda	расчётная	Anaconda Enterprise	2022

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Таблица 10

Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями

Наименование специальных помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы
1	2
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации</i>	<ol style="list-style-type: none"> 1. Экран с электроприводом 1 шт. (Инв. №558771/2) 2. Проектор 1 шт. (без инв. №) – приобретался не за счет средств вуза 3. Вандалоустойчивый шкаф 1 шт. (Инв. №558850/7) 4. Системный блок с монитором 1 шт. (Инв. №558777/9) 5. Стенд «Сергеев Сергей Степанович 1910-1999» 1 шт. (Инв. №591013/25) 6. Огнетушитель порошковый 1 шт. (Инв. №559527) 7. Подвесное крепление к огнетушителю 1 шт. (Инв. №559528) 8. Жалюзи 2шт. (Инв. №1107-221225, Инв. №1107-221225) 9. Лавка 20 шт. 10. Стол аудиторный 20 шт. 11. Стол для преподавателя 1 шт. 12. Стул 2 шт. 13. Доска маркерная 1 шт. 14. Трибуна напольная 1 шт. (без инв. №)
<i>учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы</i>	<ol style="list-style-type: none"> 1. Системный блок Intel Core Intel Core i3-2100/4096Mb/500Gb/DVD-RW 10 шт. (Инв. №601997, Инв. №601998, Инв. №601999, Инв. №602000, Инв. №602001, Инв. №602002, Инв. №602003, Инв. №602004, Инв. №602005, Инв. №602006) 2. Монитор 10 шт. (без инв. №) - приобретались не за счет средств вуза 3. Шкаф 2 шт. (Инв. №594166, Инв. №594167) 4. Тумба 1 шт. (Инв. №594168) 5. Подвесное крепление к огнетушителю 1 шт. (Инв. №

	559528) 6. Огнетушитель порошковый 1 шт. (Инв. №559527) 7. Жалюзи 1 шт. (Инв.№551557) 8. Доска магнитно-маркерная 1 шт. 9. Стол 5 шт. 10. Стол компьютерный 12 шт. 11. Стул офисный 21 шт. 12. Сейф 1 шт. (без Инв.№).
<i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы</i>	1. Трибуна напольная 1 шт. (Инв.№ 599205) 2. Шкаф для документов 3 шт. (Инв.№593633, Инв.№593634, Инв.№559548/18) 3. Вешалка напольная 2 шт. (Инв.№1107-333144, Инв.№1107-333144) 4. Жалюзи 1 шт. (Инв.№591110) 5. Доска магнитно-маркерная 1 шт. 6. Стол 15 шт. 7. Скамейка 14 шт. 8. Стол эрго 1 шт. 9. Стул 2 шт.
<i>Центральная научная библиотека имени Н.И. Железнова</i>	Читальные залы библиотеки
<i>Студенческое общежитие</i>	Комната для самоподготовки

11. Методические рекомендации обучающимся по освоению дисциплины

Приступая к изучению дисциплины «Интеллектуальный анализ данных и статистика», студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке РГАУ-МСХА им. К.А. Тимирязева, получить в библиотеке рекомендованные учебники и учебно-методические пособия, завести новую тетрадь для работы с первоисточниками.

В ходе занятий вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, записывать слова и словосочетания на иностранном языке по теме занятия, задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций или касательно иностранных слов, вызывающих затруднения при переводе.

В ходе подготовки к практическим занятиям изучить основную литературу, ознакомиться с дополнительной литературой в соответствии с поставленной задачей. При этом учесть рекомендации преподавателя и требования учебной программы. Необходимо дорабатывать свой конспект, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной учебной программой.

При подготовке к зачету и экзамену повторять пройденный материал в строгом соответствии с учебной программой. Использовать конспекты и литературу, рекомендованную преподавателем. Обратит особое внимание на темы учебных занятий, пропущенных студентом по разным причинам. При необхо-

димости обратиться за консультацией и методической помощью к преподавателю.

Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан самостоятельно выполнить сообщение (презентацию), рассмотренную на практическом занятии и подготовиться по контрольным вопросам к защите работы в рамках часов консультаций.

12. Методические рекомендации преподавателям по организации обучения по дисциплине

Курс интеллектуального анализа данных и статистики должен давать не абстрактно-формальные, а прикладные знания. Данная цель может быть реализована только при условии соблюдения в учебных планах преемственности учебных дисциплин. Базовые знания для изучения методов интеллектуального анализа данных дают такие дисциплины, как экономическая теория, информатика. Освоение подходов к интеллектуальному анализу данных и статистике позволит студентам в будущем использовать его в исследованиях, в планировании и прогнозировании.

Студент может провести собственное исследование на примере конкретного предприятия, представляющим его научный интерес, представить результаты в виде презентации. В случае надлежащего качества, его работа может быть заслушана на научном кружке кафедры или на студенческой научной конференции. По решению кафедры, студенты, занявшие призовые места на научных студенческих конференциях, могут освобождаться от сдачи зачета или экзамена по этой дисциплине.

Преподаватель должен указывать, в какой последовательности следует изучать материал дисциплины, обращать внимание на особенности изучения отдельных тем и разделов, помогать отбирать наиболее важные и необходимые сведения из учебных пособий, а также давать объяснения вопросам программы курса, которые обычно вызывают затруднения. При этом преподавателю необходимо учитывать следующие моменты:

1. Не следует перегружать студентов творческими заданиями.
2. Чередовать творческую работу на занятиях с заданиями во внеаудиторное время.
3. Давать студентам четкий инструктаж по выполнению самостоятельных заданий: цель задания; условия выполнения; объем; сроки; требования к оформлению.
4. Осуществлять текущий учет и контроль за самостоятельной работой.
5. Давать оценку обобщать уровень усвоения навыков самостоятельной, творческой работы.

Программу разработал (и):

Харитонов А.Е., к.э.н., доцент



(подпись)

РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.ДВ.06.02 «Интеллектуальный анализ данных и статистика»

ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»
(квалификация выпускника – бакалавр)

Коломеева Елена Сергеевна, доцент кафедры финансов ФГБОУ ВО г. Москвы «РГАУ-МСХА имени К.А. Тимирязева», кандидатом экономических наук (далее по тексту рецензент), проведена рецензия рабочей программы дисциплины «Интеллектуальный анализ данных и статистика» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)» (бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчик – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и кибернетики).

Рассмотрев представленные на рецензирование материалы, рецензент пришел к следующим выводам:

1. Предъявленная рабочая программа дисциплины «Интеллектуальный анализ данных и статистика» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.03.02 «Информационные системы и технологии». Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

1. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к дисциплинам по выбору части, формируемой участниками образовательных отношений учебного цикла – Б1.В.ДВ.

2. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

3. В соответствии с Программой за дисциплиной «Интеллектуальный анализ данных и статистика» закреплено 3 компетенции (7 индикаторов). Дисциплина «Интеллектуальный анализ данных и статистика» и представленная Программа способна реализовать их в объявленных требованиях.

4. Общая трудоёмкость дисциплины «Интеллектуальный анализ данных и статистика» составляет 5 зачётных единиц (18 часов/из них практическая подготовка 8 ч.).

5. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Интеллектуальный анализ данных и статистика» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии» и возможность дублирования в содержании отсутствует.

6. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

7. Программа дисциплины «Интеллектуальный анализ данных и статистика» предполагает занятия в интерактивной форме.

8. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

9. Представленные и описанные в Программе формы текущей оценки знаний (опрос, как в форме обсуждения отдельных вопросов, так и участие в деловых играх), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета, что соответствует статусу дисциплины, как факультатив-

ной дисциплины учебного цикла – ФТД ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

10. Формы оценки знаний, представленные в Программе, соответствуют специфике дисциплины и требованиям к выпускникам.

11. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 5 источников (базовый учебник), дополнительной литературой – 5 наименований, Интернет-ресурсы – 10 источников и соответствует требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

12. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Интеллектуальный анализ данных и статистика» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

13. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Интеллектуальный анализ данных и статистика».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Интеллектуальный анализ данных и статистика» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность **«Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»** (квалификация выпускника – бакалавр), разработанная Харитоновой А. Е., к.э.н., доцентом кафедры статистики и кибернетики, соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент: Коломеева Е.С., доцент кафедры финансов ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», кандидат экономических наук _____ « 26 » _____ 2022 г.

(подпись)