

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Хоружий Людмила Ивановна
Должность: Директор института экономики и управления АПК
Дата подписания: 15.07.2023 19:07:04
Уникальный программный ключ:
1e90b132d9b04dce67585160b015dddf2cb1e6a9

УТВЕРЖДАЮ:
Директор института
экономики и управления АПК
Л.И. Хоружий
« 15 » августа 2021 г.



Лист актуализации рабочей программы дисциплины Б1.О.13 Большие данные

для подготовки бакалавров
Направление 09.03.02 «Информационные системы и технологии»
Направленность: «Информационные технологии анализа данных»
Форма обучения очная
Год начала подготовки: 2019
Курс 4
Семестр 8

В рабочую программу на 2021 год начала подготовки вносятся следующие изменения:

1) Заменить в шифр Б1.О.16 на шифр Б1.О.13.

Программа актуализирована для 2021 г. начала подготовки.

Разработчик (и): Харитонов А.Е., канд. экон. наук, доцент
(ФИО, ученая степень, ученое звание)

«25» августа 2021 г.

Рабочая программа пересмотрена и одобрена на заседании кафедры статистики и кибернетики. Протокол № 11 от «26» августа 2021 г.

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

Лист актуализации принят на хранение:

Заведующий выпускающей кафедрой прикладной информатики

Худякова Е.В., д-р экон. наук, профессор Е.В. Худякова «30» августа 2021 г.



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 эконометрики

УТВЕРЖДАЮ:
Директор института экономики и управ-
ления АПК 
В.В. Бутырин
"13"  2020 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.О.16 Большие данные**

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»

Направленность: Информационные технологии анализа данных

Курс 4

Семестр 8

Форма обучения очная

Год начала подготовки 2019


Регистрационный номер _____

Москва, 2020

Разработчик (и): Харитонов А.Е., к.э.н.
(ФИО, ученая степень, ученое звание)

«16» 12 2019 г.


Рецензент: Шереужева М.А., к.э.н., доцент
(ФИО, ученая степень, ученое звание)


(подпись)
«15» 12 2019 г.

Программа составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.02 «Информационные системы и технологии» и учебного плана 2019 года начала подготовки

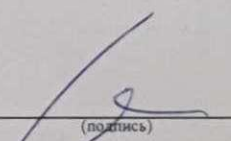
Программа обсуждена на заседании кафедры статистики и эконометрики протокол № 4 от 25 декабря 2019 г.

Зав. кафедрой Уколова А.В., к.э.н., доцент
(ФИО, ученая степень, ученое звание)



(подпись)
«15» декабря 2019 г.

Согласовано:

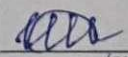
Председатель учебно-методической
комиссии института экономики и управления АПК
Корольков А.Ф., к.э.н., профессор
(ФИО, ученая степень, ученое звание)


(подпись)
«23» 01 2020 г.

Заведующий выпускающей кафедрой
прикладной информатики
Худякова Е.В., д.э.н., профессор
(ФИО, ученая степень, ученое звание)


(подпись)
«13» 01 2020 г.

Заведующий отделом комплектования ЦНБ


(подпись) Иванова Л.Л.

Бумажный экземпляр РПД, копии электронных вариантов РПД и оценочных материалов получены:

Методический отдел УМУ _____ «__» _____ 201_г

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 ТРУДОЕМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ	5
ПО СЕМЕСТРАМ	5
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	9
4.3 ЛЕКЦИИ/ ПРАКТИЧЕСКИЕ ЗАНИЯТИЯ.....	10
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	12
6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ	12
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ	12
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ	14
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	15
7.1 Основная литература	15
7.2 Дополнительная литература.....	15
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	16
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ	16
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....	18
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	18
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ.....	18

Аннотация

рабочей программы по дисциплине Б1.О.16 « Большие данные» для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности «Информационные технологии анализа данных»

Цель освоения дисциплины: Целью дисциплины «Большие данные» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области использования современных информационных технологии и программных средств, при решении задач профессиональной деятельности, исходя из имеющихся ресурсов и ограничений, что позволит эффективнее управлять своим временем.

Место дисциплины в учебном плане: дисциплина включена в обязательную часть 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-2.1; УК-2.2; УК-2.3; УК-6.1; УК-6.2; УК-6.3; ОПК-2.1; ОПК-2.2; ОПК-2.3**

Краткое содержание дисциплины:

Большие данные (Big Data): современные подходы к обработке и хранению. Терминология, история появления. Технические сложности работы с большими данными. Процесс анализа. Понятие, сущность и ключевые признаки больших данных. Роль и место больших данных в решении аналитических и исследовательских задач профессиональной деятельности. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных. Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. Программа Statistica. Язык программирования R. Другие программы. Реляционные базы данных. Параллельные базы данных. Распределённые файловые системы. NoSQL СУБД. Технология GOOGLE BIGTABLE. MapReduce. Полнотекстовый поиск. Параллельные запросы. Технология поиска и интеграции. Программные средства. ETL процесс по обработке отчётов. Понятие о технологии Data Mining. Реализация в пакетах прикладных программ. Сетевые технологии Data Mining. Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. Ассоциативные правила. Поддержка и достоверность ассоциативного правила. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Определение кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Проблемы алгоритмов кластеризации. Применение классификации и регрессии. Обзор методов классификации и регрессии.

Статистические методы. Методы, основанные на обучении, разнообразие подходов. Основные понятия теории нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура. Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле.

Общая трудоемкость дисциплины составляет: 4 зачетные единицы (144 часа).

Промежуточный контроль: экзамен, защита курсового проекта.

1. Цель освоения дисциплины

Целью дисциплины «Большие данные» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области использования современных информационных технологий и программных средств, при решении задач профессиональной деятельности, исходя из имеющихся ресурсов и ограничений, что позволит эффективнее управлять своим временем.

2. Место дисциплины в учебном процессе

Дисциплина «Большие данные» включена в обязательную часть дисциплин учебного плана. Дисциплина « Большие данные» реализуется в соответствии с требованиями ФГОС ВО, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Предшествующими курсами, на которых непосредственно базируется дисциплина « Большие данные» являются «Анализ экономических данных с использованием современных информационных технологий на иностранном языке», «Интеллектуальный анализ данных и статистика», «Расчетно-графическая статистика в среде R», «Математическая статистика», «Теория вероятностей», «Статистика».

Рабочая программа дисциплины «Большие данные» 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

Изучение данной учебной дисциплины направлено на формирование у обучающихся компетенций, представленных в таблице 1.

4. Структура и содержание дисциплины

4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 4 зач.ед. (144 часа), их распределение по видам работ семестрам представлено в таблице 2.

Таблица 1

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-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Знать: виды ресурсов и ограничений для решения профессиональных задач; основные методы оценки разных способов решения задач; действующее законодательство и правовые нормы, регулирующие профессиональную деятельность	основные принципы поиска, сбора и хранения больших данных.		
			УК-2.2 Уметь: проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения; анализировать альтернативные варианты для достижения намеченных результатов; использовать нормативно-правовую документацию в сфере профессиональной деятельности		проводить анализ поставленной цели и формулировать задачи в рамках курсового проектирования	
			УК-2.3 Владеть: методиками разработки цели и задач проекта; методами оценки потребности в ресурсах, продолжительности и стоимости проекта; навыками работы с нормативно-правовой доку-			методиками разработки цели и задач курсового проекта

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
2	УК-6	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.1 Знать: основные приемы эффективного управления собственным временем; основные методики самоконтроля, саморазвития и самообразования на протяжении всей жизни	приемы эффективного управления собственным временем при работе над курсовым проектом		
			УК-6.2 Уметь: эффективно планировать и контролировать собственное время; использовать методы саморегуляции, саморазвития и самообучения		эффективно планировать и контролировать собственное время при написании курсового проекта	
			УК-6.3 Владеть: методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний, умений и навыков; методиками саморазвития и самообразования в течение всей жизни			технологиями использования профессиональных знаний, умений и навыков в сфере обработки больших данных
	ОПК-2	Способен использовать современные информационные технологии и программные средства, в том числе отечест-	ОПК-2.1 Знать: современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной дея-	современные информационные технологии и программные средства для хранения и работы с большими данными		

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
		венного производства, при решении задач профессиональной деятельности;	тельности			
			ОПК-2.2 Уметь: выбирать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности		выбирать современные информационные технологии и программные средства для хранения и работы с большими данными	
			ОПК-2.3 Иметь навыки: применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности			присамами и навыками современных информационных технологий и программных средств, осуществляющих работу с большими данными

Таблица 2

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость, 8 семестр час.
Общая трудоёмкость дисциплины по учебному плану	144
1. Контактная работа:	53,4
Аудиторная работа	53,4
<i>в том числе:</i>	
<i>лекции (Л)</i>	24
<i>практические занятия (ПЗ)</i>	24
<i>курсовой проект (КП) (консультация, защита)</i>	3
<i>консультации перед экзаменом</i>	2
<i>контактная работа на промежуточном контроле (КРА)</i>	0,4
2. Самостоятельная работа (СРС)	90,6
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям и т.д.)</i>	26
<i>курсовой проект (КП) (подготовка)</i>	40
<i>Подготовка к экзамену</i>	24,6
Вид промежуточного контроля:	экзамен

4.2 Содержание дисциплины

Таблица 3

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнено)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ	ПКР	
Раздел 1 «Большие данные (Big Data): современные подходы к обработке и хранению»	32	6	6		20
Раздел 2 «Программное обеспечение в области анализа больших данных»	38,6	4	4		20,6
Раздел 3 «Технологии KDD и Data Mining»	78	14	14		50
<i>курсовой проект (КП) (консультация, защита)</i>	3			3	
<i>консультации перед экзаменом</i>	2			2	
Контактная работа на промежуточном контроле (КРА)	0,4			0,4	
Итого по дисциплине	144	24	24	5,4	90,6

Раздел 1 «Большие данные (Big Data): современные подходы к обработке и хранению»

Большие данные (Big Data): современные подходы к обработке и хранению. Терминология, история появления. Технические сложности работы с большими данными. Процесс анализа. Понятие, сущность и ключевые признаки больших данных. Роль и место больших данных в решении аналитических и исследовательских задач профессиональной деятельности. Общая схема анали-

за. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.

Раздел 2 «Программное обеспечение в области анализа больших данных»

Программное обеспечение в области анализа больших данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. Программа Statistica. SAS. Язык программирования R. Другие программы. Реляционные базы данных. Параллельные базы данных. Распределённые файловые системы. NoSQL СУБД. Технология GOOGLE BIGTABLE. MapReduce. Полнотекстовый поиск. Параллельные запросы. Технология поиска и интеграции. Программные средства. ETL процесс по обработке отчётов.

Раздел 3 «Технологии KDD и Data Mining»

Технологии KDD и Data Mining. Понятие о технологии Data Mining. Реализация в пакетах прикладных программ. Сетевые технологии Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. Ассоциативные правила. Поддержка и достоверность ассоциативного правила. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Определение кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Проблемы алгоритмов кластеризации. Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов. Основные понятия теории нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура. Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле.

4.3 Лекции/ практические занятия

Таблица 4

Содержание лекций, практических занятий и контрольные мероприятия

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов
1	Раздел 1 «Большие данные (Big Data): современные подходы к обработке и хранению»	Лекция № 1. Общее понятие и признаки Big Data	УК-2.1, ОПК-2.1		2
		Лекция № 2. Подходы к обработке и хранению Big Data	УК-2.2, УК-2.3, УК-6.1, УК-6.2, УК-6.3		4
		Практическая работа № 1. Поиск и определение Big Data	УК-2.1, УК-6.1, УК-6.2,	Защита работы	2

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов
			УК-6.3		
		Практическая работа № 2. Хранение больших данных	УК-2.1, ОПК-2.2, ОПК-2.3	Защита работы	4
	Раздел 2 «Программное обеспечение в области анализа больших данных»	Лекция № 3. Программное обеспечение в области анализа больших данных	УК-2.2, ОПК-2.1, ОПК-2.2		4
		Практическая работа № 3. Аналитические платформы: классификация и особенности применения	ОПК-2.1, ОПК-2.2, ОПК-2.3	Защита работы	4
	Раздел 3 «Технологии KDD и Data Mining»	Лекция № 4. Технологии KDD	УК-2.2, УК-2.3, ОПК-2.1		4
		Лекция № 5. Технологии Data Mining	УК-2.2, УК-2.3, ОПК-2.1		6
		Лекция № 6. Методология Big Data	УК-2.1, УК-2.3, ОПК-2.1		4
		Практическая работа № 4. Прогнозирование с помощью линейной регрессии	ОПК-2.2, ОПК-2.3	Защита работы	2
		Практическая работа № 5. Алгоритма кластеризации больших данных	ОПК-2.2, ОПК-2.3	Защита работы	2
		Практическая работа № 6. Поиск ассоциативных правил	ОПК-2.2, ОПК-2.3	Защита работы	4
		Практическая работа № 7. Классификация с помощью нейросети.	ОПК-2.2, ОПК-2.3	Защита работы	4
		Практическая работа № 8. Классификация с помощью деревьев решений.	ОПК-2.2, ОПК-2.3	Защита работы	2

Таблица 5

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Раздел 1 «Большие данные (Big Data): современные подходы к обработке и хранению»	Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных. (УК-2.1, УК-2.2, УК-2.3, УК-6.1, УК-6.2, УК-6.3)
2.	Раздел 2 «Программное обеспечение в области анализа больших данных»	Полнотекстовый поиск. Параллельные запросы. Технология поиска и интеграции. Программные средства. ETL процесс по обработке отчётов. (ОПК-2.1, ОПК-2.2, ОПК-2.3)
3	Раздел 3 «Технологии	Сетевые технологии Data Mining. Подготовка данных к ана-

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
	KDD и Data Mining»	лизу. Методика извлечения знаний. Data Mining. (ОПК-2.1, ОПК-2.2, ОПК-2.3)

5. Образовательные технологии

Таблица 6

Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия		Наименование используемых активных и интерактивных образовательных технологий
1.	Программное обеспечение в области анализа больших данных	Л	Лекция-визуализация
2.	Практическая работа № 5. Алгоритма кластеризации больших данных	ПЗ	Деловая игра

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

Примерная тематика курсовых проектов по дисциплине «Большие данные»:

1. Разработка хранилища данных результатов Всероссийской сельскохозяйственной переписи (другой области)
2. Применение методов машинного обучения к большим данным
3. Применение методов Data Mining к большим данным
4. Применение технологии KDD к большим данным
5. Разработка архитектуры и прототипа информационной системы обработки больших данных.
6. Применение методов классификации к большим данным
7. Применение нейронных сетей к анализу больших данных
8. Поиск ассоциативных правил в больших данных
9. Исследование и визуализация структуры Facebook, Вконтакте (других социальных сетей)
10. Social media mining
11. Программные средства анализа «больших данных» с открытым исходным кодом

Вопросы к экзамену

1. Определите сущность понятия «большие данные».
2. Опишите методики анализа больших данных.
3. Процесс аналитики анализа больших данных.

4. Особенности хранения больших данных.
5. Дайте характеристику Big Data на мировом рынке.
6. Охарактеризуйте Big Data в России.
7. Определите понятие Data Mining.
8. Определите понятие KDD.
9. Вопросы безопасности больших данных.
10. В чем состоит когнитивный анализ данных.
11. Какие модели данных вы знаете?
12. Основные описательные статистики.
13. Особенности применения корреляционно-регрессионного анализа больших данных.
14. Сущность кластерного анализа. Применение к большим данным.
15. Поиск ассоциативных правил в больших данных.
16. Классификация с помощью нейросети.
17. Классификация с помощью деревьев решений.
18. Программные средства анализа больших данных: Statistica, SPSS, Excel; их преимущества и недостатки.
19. Основные возможности хранения больших данных в языке программирования R
20. Основные возможности языка программирования R к анализу больших данных
21. Основные возможности хранения больших данных в языке программирования Python
22. Основные возможности языка программирования Python к анализу больших данных
- 23.

Пример работ

Практическая работа № 1. Поиск и определение Big Data.

Требуется осуществить поиск источников информации в сети Интернет: открытые и закрытые источники данных. Рассмотреть портал открытых данных РФ. Определить можно ли классифицировать найденные данными как «Big Data». Сохранить данные. Написать отчет с выводами.

Практическая работа № 2. Хранение больших данных

Описать модель организации хранилища данных. Загрузить данные в RStudio. Организовать связь между загруженными таблицами с помощью SQL-запросов. Осуществить выгрузку информации для просмотра описательных статистик. Написать отчет с выводами.

Практическая работа № 3. Аналитические платформы: классификация и особенности применения

Организовать импорт данных из работы 2 из RStudio в пакет Statistica. Описать структуру полученных данных. Провести описательную статистику в пакете Statistica. Написать отчет с выводами.

Практическая работа № 4. Прогнозирование с помощью линейной регрессии.

Построить уравнение линейной регрессии по большим данным, загруженным в RStudio во 2 работе средствами языка программирования R. Написать отчет с выводами.

Практическая работа № 5. Алгоритма кластеризации больших данных

Провести кластерный Большие данные, загруженных в RStudio во 2 работе. Применить метод k-средних и иерархическую классификацию. Выбрать оптимальный способ. Написать отчет с выводами.

Практическая работа № 6. Поиск ассоциативных правил

По показателям больших данных, загруженных в RStudio во 2 работе, осуществить поиск ассоциативных правил. Написать отчет с выводами.

Практическая работа № 7. Классификация с помощью нейросети.

Методами нейросетевой классификации определить возможность разбиения больших данных на кластеры, выделенные в работе 5 по показателям, характеризующим другую сферу явления. Написать отчет с выводами.

Практическая работа № 8. Классификация с помощью деревьев решений.

К показателям больших данных, загруженных в RStudio во 2 работе, применить алгоритм «Дерево решений». Описать структуру дерева. Выявить основные факторы, влияющие на классификацию. Написать отчет с выводами.

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

Текущий контроль знаний, умений и навыков проводится в форме теоретических вопросов, 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.

Оценка знаний ведется на основе рейтинговой оценки студента, которая складывается из средней оценки за выполнение индивидуальных заданий на практических занятиях. Максимальная оценка за выполнение индивидуального задания - 10 баллов. Оценка 9 ставится при наличии нарушений норм в оформлении работы. Оценка 8 – при наличии негрубых вычислительных ошибок, которые не привели к ложным выводам и неверному пониманию сути работы. Оценка 7 – сделаны неверные выводы вследствие ошибки в расчетах, при этом не нарушена логика исследования. Оценка 6-5– нарушена логика анализа, ошибочные выводы. Задержка выполнения индивидуального практического задания на одну неделю штрафуются одним баллом, на две - двумя. По истечении трех недель с момента выполнения задания в аудитории работа не принимается, а лекция по данной теме становится дополнительной темой беседы на экзамене.

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. Методы и модели исследования сложных систем и обработки больших данных : монография / И.Ю. Парамонов, В.А. Смагин, Н.Е. Косых, А.Д. Хомоненко ; под редакцией В.А. Смагина и А.Д. Хомоненко. – Санкт-Петербург : Лань, 2020. – 236 с. – ISBN 978-5-8114-4006-1. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/126938> (дата обращения: 14.02.2020). – Режим доступа: для авториз. пользователей.
2. Макшанов, А.В. Технологии интеллектуального анализа данных : учебное пособие / А.В. Макшанов, А.Е. Журавлев. – 2-е изд., стер. – Санкт-Петербург : Лань, 2019. – 212 с. – ISBN 978-5-8114-4493-9. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/120063> (дата обращения: 14.02.2020). – Режим доступа: для авториз. пользователей.
3. Нестеров, С.А. Основы интеллектуального анализа данных. Лабораторный практикум : учебное пособие / С.А. Нестеров. – Санкт-Петербург : Лань, 2020. – 40 с. – ISBN 978-5-8114-4509-7. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/130181> (дата обращения: 14.02.2020). – Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Трушков, А.С. Статистическая обработка информации. Основы теории и компьютерный практикум + CD : учебное пособие / А.С. Трушков. – Санкт-Петербург : Лань, 2020. – 152 с. – ISBN 978-5-8114-4322-2. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/126947> (дата обращения: 14.02.2020). – Режим доступа: для авториз. пользователей.
2. Буховец, А.Г. Алгоритмы вычислительной статистики в системе R : учебное пособие / А.Г. Буховец, П.В. Москалев. – 2-е изд., перераб. и доп. – Санкт-Петербург : Лань, 2015. – 160 с. – ISBN 978-5-8114-1802-2. – Текст : электронный // Электронно-библиотечная система «Лань» : [сайт]. – URL: <https://e.lanbook.com/book/68459>(открытый доступ)
3. Абдикеев, Н.М. Когнитивная бизнес-аналитика: учебное пособие / Н.М. Абдикеев, А.Н. Аверкин, Л.П. Дьяконова и др. – М.: Инфра-М, 2014. – 509 с.
4. Образцова О.И. Статистика предприятий и бизнес-статистика [Текст] : учебное пособие для студентов высших учебных заведений по направлению подготовки "Экономика" / О.И. Образцова. - Москва : Высшая школа экономики, 2011. – 698с.

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

1. Bureau of Economic Analysis. URL: <http://www.bea.gov> (открытый доступ)
2. Econometric Laboratory Software Archive. URL: <http://elsa.berkeley.edu/> (открытый доступ)
3. NASS – National Agricultural Statistics Service. URL: www.nass.usda.gov (открытый доступ)
4. STATISTICS. URL: <http://www.oecd-ilibrary.org/statistics;jsessionid=3ddci6tti4o90.delta> (открытый доступ)
5. STATISTICS. URL: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes> (открытый доступ)
6. Доклады о развитии человека. URL: <http://www.un.org/ru/development/hdr/> (открытый доступ)
7. Московская международная валютная биржа. <http://www.micex.ru> (открытый доступ)
8. Основные обзоры и доклады ООН в экономической и социальной областях. URL: <http://www.un.org/ru/development/surveys/> (открытый доступ)
9. Официальный сайт Всемирного банка . URL: <http://www.worldbank.org> (открытый доступ)
10. Официальный сайт Международного валютного фонда. URL: <http://www.imf.org> (открытый доступ)
11. Официальный сайт Национального бюро статистики по рынку труда США. URL: <http://www.bls.gov> (открытый доступ)
12. Официальный сайт Национального бюро экономических исследований США. URL: <http://www.nber.org> (открытый доступ)
13. Официальный сайт Росстата. URL: <http://www.gks.ru/> (открытый доступ)
14. Официальный сайт Центрального Банка России. URL: (открытый доступ) <http://www.cbr.ru>
15. Центр макроэкономического анализа и прогнозирования при ИНП РАН. <http://www.forecast.ru> (открытый доступ)
16. The R Project for Statistical Computing <https://www.r-project.org/> (открытый доступ)
17. Язык программирования Python. URL: <https://www.python.org/> (открытый доступ)
18. Анаконда. URL: <https://www.anaconda.com/distribution/> (открытый доступ)
19. Банк данных для анализа. URL: <https://www.kaggle.com/> (открытый доступ)

9. Перечень программного обеспечения

Таблица 9

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	1-3	R	расчётная	r-project	2018

2	1-3	RStudio	расчётная	r-project	2018
3	2	Statistica	расчетная	StatSoft	2004
4	1-3	Anaconda	расчётная	Anaconda Enterprise	2019

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, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации (02-102)</i>	Мультимедиа
<i>Учебная аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, курсового проектирования, текущего контроля и промежуточной аттестации (02-302)</i>	10 компьютеров конфигурации: INTEL Core i3-2100/4096 Mb/500Gb/DVD-RW, MS Word, MS Excel, пакеты прикладных программ: STATA, R, Python, Eviews, Statistica, доступ к сети Internet, справочной правовой системе КонсультантПлюс
Студенческое общежитие	Комнаты для самоподготовки
ЦНБ имени Н.И. Железнова	Читальный зал

11. Методические рекомендации студентам по освоению дисциплины

Приступая к изучению дисциплины « Большие данные», 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.

Курс предусматривает, что студенты обладают необходимым уровнем знаний по информатике, математике, экономике, приходят на практические занятия подготовленными по вопросам лекционного материала. Предполагается, что студент выполняет практическое задание в аудитории, дома оформляет и готовится по теоретическим вопросам к защите работы на следующем занятии.

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

Комплексное освоение студентами учебной дисциплины «Большие данные» предполагает изучение материалов лекций, рекомендуемой учебно-методической литературы, подготовку к практическим занятиям и лекциям, са-

циям, самостоятельную работу при выполнении практических заданий, домашних заданий.

На первом занятии преподаватель закрепляет за каждым студентом номер варианта для выполнения индивидуальных работ (как правило, номер варианта соответствует порядковому номеру студента в журнале преподавателя). По каждой индивидуальной работе должна быть поставлена оценка по факту ее защиты. Защиту рекомендуется проводить на следующем после получения задания занятии. Преподаватель обязан проверить соответствие выполненного задания исходным данным варианта студента. Таким образом, исключается вероятность плагиата.

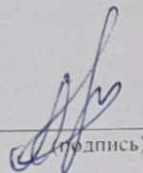
В рамках курса предусмотрены формы работы студентами в малых группах, мозговой штурм, ориентированные на развитие навыков взаимодействия у студентов при решении профессиональных задач.

Преподаватель должен стимулировать студентов к занятию научно-исследовательской работой, изучению научной эконометрической литературы, в т.ч. отечественной и зарубежной периодики.

Студент может провести собственное статистическое наблюдение за социально-экономическими явлениями, представляющими его научный интерес, построить статистическую модель, сделать прогноз. 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 освободиться от сдачи экзамена по дисциплине.

Программу разработал (и):

Харитонов А.Е., к.э.н., доцент



(подпись)

РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.О.16 «**Большие данные**»
ОПОП ВО по направлению 09.03.02 «*Информационные системы и технологии*», направ-
ленность «**Информационные технологии анализа данных**» (квалификация выпускни-
ка – бакалавр)

Шереужева Маджиной Альбертовной, доцентом кафедры финансов ФГБОУ ВО г. Москвы «РГАУ-МСХА имени К.А. Тимирязева», кандидатом экономических наук (далее по тексту рецензент), проведено рецензирование рабочей программы дисциплины «**Большие данные**» ОПОП ВО по направлению 09.03.02 «*Информационные системы и технологии*», направленность Информационные технологии анализа данных (уровень обучения - бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и эконометрики (разработчик – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и эконометрики).

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

1. Предъявленная рабочая программа дисциплины «**Большие данные**» (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.

2. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к дисциплинам обязательной части учебного цикла – Б1.О.

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

4. В соответствии с Программой за дисциплиной «**Большие данные**» закреплено 3 компетенции (9 индикаторов). Дисциплина «**Большие данные**» и представленная Программа способна реализовать их в объявленных требованиях.

5. Результаты обучения, представленные в Программе в категориях знать, уметь, владеть соответствуют специфике и содержанию дисциплины и демонстрируют возможность получения заявленных результатов.

6. Общая трудоёмкость дисциплины «**Большие данные**» составляет 4 зачётные единицы (144 часа).

7. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «**Большие данные**» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.02 «*Информационные системы и технологии*» и возможность дублирования в содержании отсутствует.

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

9. Программа дисциплины «**Большие данные**» предполагает проведение занятий в интерактивной форме.

10. 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*».

11. Представленные и описанные в Программе формы текущей оценки знаний (защита работ), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме экзамена и защиты курсового проекта в 8 семестре, что соответствует статусу дисциплины, как дисциплины обязательной части учебного цикла – Б1.О ФГОС ВО направления 09.03.02 «*Информационные системы и технологии*».

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

13. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 3 источника (базовый учебник), дополнительной литературой – 4 наименования Интернет-ресурсы – 19 источников и соответствует требованиям ФГОС ВО направления 09.03.02 «*Информационные системы и технологии*».

14. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «**Большие данные**» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

15. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «**Большие данные**».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «**Большие данные**» ОПОП ВО по направлению 09.03.02 «*Информационные системы и технологии*», направленность «*Информационные технологии анализа данных*» (квалификация выпускника – бакалавр), разработанная Харитоновой А. Е., к.э.н., доцентом кафедры статистики и эконометрики, соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент: Шереужева М.А., доцент кафедры финансов ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», кандидат экономических наук _____ « 15 » 12

(подпись)

2019 г.