

Документ подписан простой электронной подписью
Информация о владельце: **МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ**
ФИО: Хоружий, Л.И. Иванов
Должность: Директор института экономики и управления АПК
Дата подписания: 12.05.2022
Уникальный идентификационный ключ:
1e90b132d9b040c267583160b015dddf2cb1e6a9



ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ 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
“ 12 ” 05 2022 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.02 Наука о данных (Data Science)**

для подготовки магистров

ФГОС ВО

Направление: 09.04.02 «Информационные системы и технологии»
Направленность: Науки о данных (Data Science)

Курс 1
Семестр 1,2

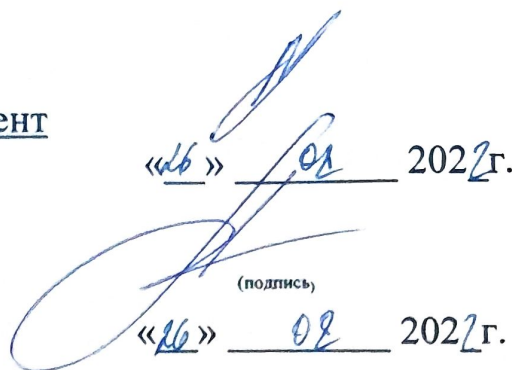
Форма обучения очная
Год начала подготовки 2022

Москва, 2022

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

(ФИО, ученая степень, ученое звание)

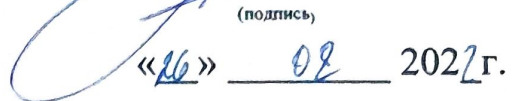
«26» 08 2022г.


(подпись)

Рецензент: Коломеева Е.С., к.э.н.

(ФИО, ученая степень, ученое звание)

«26» 08 2022г.


(подпись)

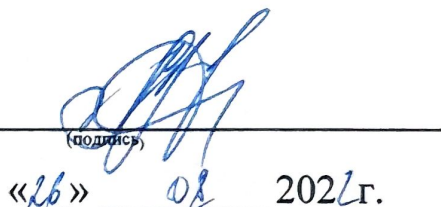
Программа составлена в соответствии с требованиями ФГОС ВО, профессионального стандарта и учебного плана по направлению подготовки 09.03.02 «Информационные системы и технологии».

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)

«26» 08 2022г.

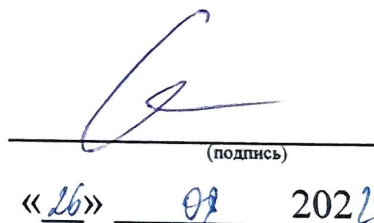

(подпись)

Согласовано:

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

(ФИО, ученая степень, ученое звание)

«26» 08 2022г.

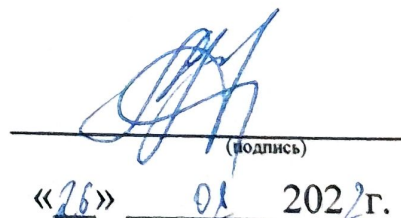

(подпись)

И.о.заведующего выпускающей кафедрой
статистики и кибернетики

Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)

«26» 08 2022г.


(подпись)

/Заведующий отделом комплектования ЦНБ


(подпись)

Содержание

АННОТАЦИЯ	4
1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ	6
2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ	6
3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	6
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	7
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ	7
ПО СЕМЕСТРАМ	7
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	10
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	12
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....	14
6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	14
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ.....	14
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ.....	17
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	18
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	18
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	18
7.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ, РЕКОМЕНДАЦИИ И ДРУГИЕ МАТЕРИАЛЫ К ЗАНЯТИЯМ	19
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	19
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ.....	19
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....	20
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ..	20
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	21

Аннотация
рабочей программы учебной дисциплины
Б1.В.02 «Наука о данных (Data Science)»

для подготовки магистров по направлению 09.04.02 «Информационные системы и технологии» по направленности Науки о данных (Data Science)

и

Цель освоения дисциплины: освоение студентами теоретических и практических знаний и приобретение умений и навыков в области разработки оригинальных алгоритмов и программных средств, анализа профессиональной информации и применения на практике новых научных принципов и методов исследования для решения профессиональных задач.

Место дисциплины в учебном плане: включена в часть, формируемую участниками образовательных отношений учебного плана по направлению подготовки 09.04.02 «Информационные системы и технологии».

Требования к результатам освоения дисциплины: в результате освоения дисциплины формируются следующие компетенции (индикаторы): ПКос-2 (ПКос-2.1; ПКос-2.2; ПКос-2.3); ПКос-3 (ПКос-3.1, ПКос-3.2; ПКос-3.3).

Краткое содержание дисциплины:

Основные этапы жизненного цикла данных. Создание данных (Data Generation/Data Capture). Обслуживание данных (Data Maintenance). Синтез данных (Data Synthesis). Использование данных (Data Usage). Публикация данных (Data Publication). Архивация данных (Data Archival). Уничтожение данных (Data Purging). Линейная алгебра. Векторы. Матрицы. Статистика. Описание одиночного набора данных. Показатели центральной тенденции. Показатели вариации. Корреляция. Парадокс Симпсона. Теория вероятностей. Зависимость и независимость. Условная вероятность. Теорема Байеса. Случайные величины. Непрерывные распределения. Нормальное распределение. Центральная предельная теорема. Проверка статистических гипотез. Доверительные интервалы. Байесовский статистический вывод. Градиентный спуск, методы оценивания градиента. Функции потерь. Общая характеристика программных средств обработки данных. Программы Business Intelligence. Язык программирования Python: библиотеки для машинного обучения Matplotlib, Pandas, Scikit-learn, Mlglearn, SciPy, Numpy, Keras. Язык программирования R: библиотеки для машинного обучения Pandas, Dplyr, [Ggplot2](#), [MLR](#), [Caret](#), [Esquisse](#), H2O. Постановки основных классов задач в машинном обучении. Обучение с учителем: регрессия и классификация. Обучение без учителя: кластеризация, снижение размерности. Рекомендательные системы, обработка текстов: тематическое моделирование, построение аннотаций, извлечение ответов на вопросы, машинный перевод. Обучение с подкреплением. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. Методы отбора признаков. Метод главных компонент и singular spectrum analysis. Ядровые

методы. Ядра и спрямляющие пространства, методы их построения. Операции в спрямляющих пространствах. Ограничения линейных методов. Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес. Градиентный бустинг. Нейронные сети: общая архитектура. Теорема об универсальной аппроксимации. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. Нейронные сети в обработке изображений. Фильтры. Сверточные слои. Нейронные сети и обучение представлений. Обработка последовательностей. Рекуррентные нейронные сети. **Общая трудоемкость дисциплины составляет:** составляет 2 зачетные единицы (72 часа).

Промежуточный контроль: с оценкой.

1. Цель освоения дисциплины

Целью дисциплины «Наука о данных (Data Science)» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области разработки оригинальных алгоритмов и программных средств, анализа профессиональной информации и применения на практике новых научных принципов и методов исследования для решения профессиональных задач.

2. Место дисциплины в учебном процессе

Дисциплина «Наука о данных (Data Science)» включена в часть, формируемую участниками образовательных отношений учебного плана. Дисциплина «Наука о данных (Data Science)» реализуется в соответствии с требованиями ФГОС ВО, профессионального стандарта, ОПОП ВО и Учебного плана по направлению 09.04.02 «Информационные системы и технологии».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Наука о данных (Data Science)» являются «Инструменты Data Science в R, Python, SQL», «Эконометрика (продвинутый уровень)».

Дисциплина «Наука о данных (Data Science)» является основополагающей для изучения следующих дисциплин: «Аналитика и визуализация данных на Python на иностранном языке», «Аналитика и визуализация данных на Python», «Глубокое обучение в науках о данных», «Системы поддержки принятия решений», «Анализ больших данных (Big Data Analytics)», «Анализ больших данных (Big Data Analytics)».

Рабочая программа дисциплины «Наука о данных (Data Science)» 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 2 зач.ед. (72 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.	ПКос-2	Способен применять современные средства анализа данных и бизнес-аналитики в профессиональной деятельности	ПКос-2.1 Знать: инструментальные средства бизнес-аналитики и пакеты прикладных статистических программ	инструментальные средства Data Science		
			ПКос-2.2 Уметь: осуществлять сбор, обработку, анализ массовых данных, строить прогнозы с использованием современных инструментов бизнес-аналитики и пакетов прикладных статистических программ		осуществлять сбор, обработку, анализ массовых данных, строить прогнозы с использованием инструментов Data Science	
			ПКос-2.3 Иметь навыки: сбора, обработки, анализа массовых данных, в т.ч. больших данных, эконометрического моделирования и прогнозирования с использованием современных средств бизнес-аналитики и пакетов прикладных статистических программ			навыками сбора, обработки, анализа массовых данных с использованием инструментов Data Science
2.	ПКос-3	Способен совершенствовать и разрабатывать	ПКос-3.1 Знать: методы науки о данных, в т.ч. методы машинного обучения,	методы науки о данных, в т.ч. методы машинного обучения, обработки и		

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
		новые методы, модели, алгоритмы, технологии и инструментальные средства работы с данными, в т.ч. большими данными	обработки и визуализации больших данных; состояние и перспективы развития науки о данных, используемого при обработке данных программного инструментария; потребности в совершенствовании и разработке новых методов, технологий и инструментальных средств для работы с данными, в т.ч. большими; область применения науки о данных в сельском хозяйстве	визуализации больших данных; состояние и перспективы развития науки о данных, используемого при обработке данных программного инструментария потребности в совершенствовании и разработке новых методов, технологий и инструментальных средств для работы с данными, в т.ч. большими; область применения науки о данных в сельском хозяйстве		
			ПКос-3.2 Уметь: определять перспективную тематику научно-исследовательских работ в области совершенствования и разработки новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с данными; планировать и проводить аналитические и научные		планировать и проводить аналитические и научные исследования по тематике информационных технологий, применяемых в науке о данных	

№ п/п	Код компете нции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
			исследования по тематике информационных технологий, применяемых в науке о данных			
			ПКос-3.3 Иметь навыки: разработки новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с данными на основе анализа потребностей и передового зарубежного и отечественного опыта; планирования состава и содержания, согласование перечня научно-исследовательских работ в профессиональной деятельности			разработки новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с данными

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость час. всего/*		
	час.	в т.ч. по семестрам	
		1	2
Общая трудоёмкость дисциплины по учебному плану	72	36	36
1. Контактная работа:	14,35/2	2	12,35/2
Аудиторная работа	14,35/2	2	12,35/2
<i>в том числе:</i>			
лекции (Л)	6	2	4
практические занятия (ПЗ)	8/2		8/2
контактная работа на промежуточном контроле (КРА)	0,35		0,35
2. Самостоятельная работа (СРС)	57,65	34	23,65
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям и т.д.)</i>	43,65	34	19,65
Подготовка к зачету с оценкой (контроль)	4		4
Вид промежуточного контроля:			зачет с оценкой

* в том числе практическая подготовка.

4.2 Содержание дисциплины

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнёно)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ/*	ПКР	
Установочная лекция	36	2			34
Итого за 1 семестр	36	2	0	0	34
Раздел 1 Математический инструментарий науки о данных.	7	2	1		4
Раздел 2 Программный инструментарий науки о данных.	7	2	1		4
Раздел 3 Машинное обучение	13,65		4/2		9,65
Раздел 4 Глубокое обучение	8		2		6
Контактная работа на промежуточном контроле (КРА)	0,35			0,35	
Итого за 2 семестр	36	4	8/2	0,35	23,65
Итого по дисциплине	72	6	8/2	0,35	57,65

* в том числе практическая подготовка

Раздел 1 Математический инструментарий науки о данных.

Основные этапы жизненного цикла данных. Создание данных (Data Generation/Data Capture). Обслуживание данных (Data Maintenance). Синтез данных (Data Synthesis). Использование данных (Data Usage). Публикация данных (Data Publication). Архивация данных (Data Archival). Уничтожение данных (Data Purging).

Линейная алгебра. Векторы. Матрицы. Статистика. Описание одиночного набора данных. Показатели центральной тенденции. Показатели вариации. Корреляция. Парадокс Симпсона. Теория вероятностей. Зависимость и независимость. Условная вероятность. Теорема Байеса. Случайные величины. Непрерывные распределения. Нормальное распределение. Центральная предельная теорема. Проверка статистических гипотез. Доверительные интервалы. Байесовский статистический вывод. Градиентный спуск, методы оценивания градиента. Функции потерь.

Раздел 2 Программный инструментарий науки о данных.

Общая характеристика программных средств обработки данных. Программы Business Intelligence. Язык программирования Python: библиотеки для машинного обучения Matplotlib, Pandas, Scikit-learn, Mgllearn, SciPy, Numpy, Keras. Язык программирования R: библиотеки для машинного обучения Pandas, Dplyr, [Ggplot2](#), [MLR](#), [Caret](#), [Esquisse](#), H2O.

Раздел 3 Машинное обучение

Постановки основных классов задач в машинном обучении. Обучение с учителем: регрессия и классификация. Обучение без учителя: кластеризация, снижение размерности. Рекомендательные системы, обработка текстов: тематическое моделирование, построение аннотаций, извлечение ответов на вопросы, машинный перевод. Обучение с подкреплением. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. Методы отбора признаков. Метод главных компонент и singular spectrum analysis. Ядровые методы. Ядра и спрямляющие пространства, методы их построения. Операции в спрямляющих пространствах. Ограничения линейных методов. Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес. Градиентный бустинг.

Раздел 4 Глубокое обучение

Нейронные сети: общая архитектура. Теорема об универсальной аппроксимации. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. Нейронные сети в обработке изображений. Фильтры. Сверточные слои. Нейронные сети и обучение представлений. Обработка последовательностей. Рекуррентные нейронные сети.

4.3 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
		Установочная лекция	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3		2
	Раздел 1 Математический инструментарий науки о данных	Лекция № 1. Математический инструментарий науки о данных	ПКос-2.1 ПКос-3.1		2
		Практическая работа № 1. Основные математические инструменты обработки данных	ПКос-2.1 ПКос-2.2 ПКос-3.1 ПКос-3.2	Вопросы дискуссии	1
3	Раздел 2 Программный инструментарий науки о данных.	Лекция № 2. Общая характеристика программных средств обработки данных	ПКос-2.1 ПКос-3.1		2
		Практическая работа № 2. Обзор зарубежных программных средств обработки данных	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	Вопросы дискуссии	1
4	Раздел 3 Машинное обучение	Практическая работа № 3. Методы машинного обучения «без учителя»	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	Защита работы	2
		Практическая работа № 4. Методы машинного обучения «с учителем»	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	Защита работы	2/2
5	Раздел 4 Глубокое обучение	Практическая работа № 5. Построение нейронной сети	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	Защита работы	0,5
		Практическая работа № 6. Классификация с помощью нейросети.	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	Защита работы	0,5

№ п/п	№ раздела	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов/ из них практическая подготовка
		Итоговая контрольная работа	ПКос-2.1 ПКос-2.2 ПКос-2.3 ПКос-3.1 ПКос-3.2 ПКос-3.3	письменная работа	1

Таблица 5

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Раздел 1 Математический инструментальный науки о данных	Пол. Теорема Байеса. Случайные величины. Непрерывные распределения. Байесовский статистический вывод. Градиентный спуск, методы оценивания градиента. Функции потерь. (ПКос-2.1, ПКос-2.2, ПКос-2.3, ПКос-3.1, ПКос-3.2, ПКос-3.3)
2	Раздел 2 Программный инструментальный науки о данных.	Язык программирования Python: библиотеки для машинного обучения Mllearn, Keras. Язык программирования R: библиотеки для машинного обучения MLR , Esquisse , H2O. (ПКос-2.1, ПКос-2.2, ПКос-2.3, ПКос-3.1, ПКос-3.2, ПКос-3.3)
3	Раздел 3 Машинное обучение	Обучение с подкреплением. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Кросс-валидация: тонкости (отбор переменных, переобучение на валидационное множество). Оценки ожидаемой ошибки для линейной регрессии: AIC и другие. L1 и L2 регуляризация. (ПКос-2.1, ПКос-2.2, ПКос-2.3, ПКос-3.1, ПКос-3.2, ПКос-3.3)
4	Раздел 4 Глубокое обучение	Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. Нейронные сети в обработке изображений. Фильтры. Сверточные слои. (ПКос-2.1, ПКос-2.2, ПКос-2.3, ПКос-3.1, ПКос-3.2, ПКос-3.3)

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.	Практическая работа № 4. Методы машинного обучения «с учителем»	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

Вопросы к зачету с оценкой

1. Определите сущность понятия «Наука о данных (Data Science)».
2. Отличия Data Science от других наук.
3. Что такое данные? Какие ГОСТы с определениями данных вам известны?
4. Что такое жизненный цикл данных? Перечислите этапы жизненного цикла данных.
5. Для каких целей нужен этап «Синтез данных»
6. Для каких целей нужен этап «Использование данных»
7. Для каких целей нужен этап «Публикация данных»
8. Для каких целей нужен этап «Архивация данных»
9. Роль математического инструментария в науке о данных.
10. Роль программного инструментария в науке о данных.
11. Методы машинного обучения «с учителем».
12. Методы машинного обучения «без учителя».
13. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
14. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
15. Метрики качества алгоритм регрессии и классификации.
16. Деревья решений. Методы построения деревьев. Их регуляризация.
17. Случайный лес, его особенности.
18. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.
19. Кластеризация. Алгоритм K-Means.
20. Классификация с помощью нейросети.
21. Из чего должна состоять простая нейронная сеть?
22. Как обучение представлений связано с глубокими нейросетями?
23. Как обучать нейронную сеть? Какие адаптивные методы оптимизации вы знаете?

Итоговая контрольная работа

Загрузить данные по своему варианту в Python. Рассчитать по показателям основные статистики (среднюю, дисперсию, коэффициент вариации, медиану). Провести классификацию «с обучением» и «без обучения» методами:

- дерево решений;
- случайный лес;

- ближайший сосед;

Примените метод опорных векторов с предварительным обучением нейронной сети. Сделайте выводы.

Пример работ

Практическая работа № 1. Основные математические инструменты обработки данных

Вопросы к дискуссии:

1. Какие математические инструменты необходимы специалисту в области data science?
2. Роль математического инструментария в науке о данных.
3. Создание и работы с векторами и матрицами в R и Python.
4. Какие описательные статистики Вы помните? Как они интерпретируются?
5. Расчет основных описательных статистик в R и Python.
6. Назовите параметрические и непараметрические статистические гипотезы. В чем их отличия?
7. Проверка статистических гипотез в R и Python.
8. Теорема Байеса: суть и применение.
9. В чем суть и практическая значимость градиентного спуска?
10. Назовите методы оценивания градиента.

Практическая работа № 2. Обзор зарубежных программных средств обработки данных

Вопросы к дискуссии:

1. Назовите наиболее популярные в России программные средства обработки данных?
2. Какие программные средства обработки данных популярны в США?
3. Найдите примерную стоимость программных средств Business Intelligence.
4. В чем особенность использования языков R и Python при обработке данных?
5. Какие программные средства Business Intelligence наиболее популярны в мире?
6. Какие на Ваш взгляд программные средства обработки данных приобретут популярность в ближайшие годы?

Практическая работа № 3. Методы машинного обучения «без учителя»

Загрузить исходные данные в RStudio и Spyder. Для данных своего варианта с помощью провести классификацию «без учителя» методами: дерево решений, случайный лес, опорные вектора, ближайший сосед, наивный Байесовский классификатор. Сравнить результаты, полученные при помощи двух разных языков программирования. Оформить отчет с выводами.

Вопросы к защите:

1. Методы машинного обучения «с учителем».
2. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.

3. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
4. Метрики качества алгоритм регрессии и классификации.
5. Деревья решений. Методы построения деревьев. Их регуляризация.
6. Случайный лес, его особенности.

Практическая работа № 4. Методы машинного обучения «с учителем»

Загрузить исходные данные в RStudio и Spyder. Для данных своего варианта с помощью провести классификацию «с учителем» методами: дерево решений, случайный лес, опорные вектора, ближайший сосед, наивный Байесовский классификатор. Сравнить результаты, полученные при помощи двух разных языков программирования. Сравнить методы «с учителем» с методами «без учителя». Оформить отчет с выводами.

Вопросы к защите:

1. Методы машинного обучения «без учителя».
2. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
3. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
4. Метрики качества алгоритм регрессии и классификации.
5. Деревья решений. Методы построения деревьев. Их регуляризация.
6. Случайный лес, его особенности.

Практическая работа № 5 . Построение нейронной сети

Загрузить исходные данные в RStudio и Spyder. Применить методов опорных векторов с применением нейросети. Использовать различные пакеты для построения нейросети. Оформить отчет с выводами.

Вопросы к защите:

1. Из чего должна состоять простая нейронная сеть?
2. Как определить оптимальное число слоев?
3. Как обучение представлений связано с глубокими нейросетями?

Практическая работа № 6. Классификация с помощью нейросети.

Загрузить исходные данные в RStudio и Spyder. Построить модель классификации на основе искусственных нейронных сетей. Сравнить результаты классификации при разном уровне скрытых слоев. Сравнить качество классификации с предыдущими методами классификации. Сравнить результаты построенных моделей. Оформить отчет с выводами.

Вопросы к защите:

1. Классификация с помощью нейросети.
2. Качество классификации
3. Проблемы переобучения

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 штрафуеться 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 зачете с оценкой 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. Data Science / Francesco Palumbo, Angela Montanari, Maurizio Vichi. Springer International Publishing AG, 2017 – Текст : электронный // Springer: электронно-библиотечная система. URL: <https://link.springer.com/book/10.1007/978-3-319-55723-6#editorsandaffiliations>
2. 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>
3. 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>

4. Токмаков, Г. П. Базы данных: Модели и структуры данных, язык SQL, программирование баз данных : учебное пособие / Г. П. Токмаков. — Ульяновск : УлГТУ, 2021. — 362 с. — ISBN 978-5-9795-2184-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/259706>

5. R for Data Science / Н. Wickham, G. Grolemund – Canada: O’Reilly Media, 2017, 494 p.

7.2 Дополнительная литература

1. 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>

2. Скляр, А. Я. Системы управления данными : учебное пособие / А. Я. Скляр, А. А. Высоцкая, А. А. Горячев. — Москва : РТУ МИРЭА, 2022. — 163 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/265730> (дата обращения: 16.12.2022). — Режим доступа: для авториз. пользователей.

3. Забелин, А. А. Реализация алгоритмов вычислительной математики на языке Python : учебное пособие / А. А. Забелин. — Чита : ЗабГУ, 2020. — 130 с. — ISBN 978-5-9293-2575-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/173632> (дата обращения: 16.12.2022). — Режим доступа: для авториз. пользователей.

4. Перекатов, А. С. Статистическая обработка экспериментальных данных. Полный факторный эксперимент в языке R : учебное пособие / А. С. Перекатов, М. Б. Никифоров. — Рязань : РГРТУ, 2019. — 48 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/168309> (дата обращения: 16.12.2022). — Режим доступа: для авториз. пользователей.

5. 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, 2022. — 160 с. — ISBN 978-5-8114-1802-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/212195> (дата обращения: 16.12.2022). — Режим доступа: для авториз. пользователей.

7.3 Методические указания, рекомендации и другие материалы к занятиям

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

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

1. Язык программирования Python. URL: <https://www.python.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: <https://rosstat.gov.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](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
	Раздел 1 Математический инструментарий науки о данных Раздел 2 Программный инструментарий науки о данных. Раздел 3 Машинное обучение Раздел 4 Глубокое обучение	R	расчётная	r-project	2022
1	Раздел 1 Математический инструментарий науки о данных Раздел 2 Программный инструментарий науки о данных. Раздел 3 Машинное обучение Раздел 4 Глубокое обучение	RStudio	расчётная	r-project	2022
2	Раздел 1 Математический инструментарий науки о данных	Anaconda	расчётная	Anaconda Enterprise	202

Раздел 2 Программный инструментарий науки о данных. Раздел 3 Машинное обучение Раздел 4 Глубокое обучение				
--	--	--	--	--

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 i3-2100/4096Мб/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</i>	<ol style="list-style-type: none"> 1. Трибуна напольная 1 шт. (Инв.№ 599205) 2. Шкаф для документов 3 шт. (Инв.№593633, Инв.№593634, Инв.№559548/18)

<i>индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации, помещение для самостоятельной работы</i>	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 «Наука о данных (Data Science)», 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 консул

12. Методические рекомендации преподавателям по организации обучения по дисциплине

Комплексное освоение студентами учебной дисциплины «Наука о данных (Data Science)» предполагает изучение материалов лекций, рекомендуемой учебно-методической литературы, подготовку к практическим занятиям и лекциям, 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.

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 освобождаться от сдачи экзамена по дисциплине.

Программу разработал (и):

Харитонов А.Е., к.э.н., доцент



(подпись)

РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.02 «Наука о данных (Data Science)»
ОПОП ВО по направлению 09.04.02 «Информационные системы и технологии»,
направленность «Науки о данных (Data Science)»
(квалификация выпускника – магистр)

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 «Наука о данных (Data Science)» ОПОП ВО по направлению 09.04.02 «Информационные системы и технологии», направленность «Науки о данных (Data Science)» (магистриат) 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 «Наука о данных (Data Science)» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.04.02 «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 части, формируемой участниками образовательных отношений учебного цикла – Б1.В.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 09.04.02 «Информационные системы и технологии».

4. В соответствии с Программой за дисциплиной «Наука о данных (Data Science)» закреплено 2 компетенции (6 индикаторов). Дисциплина «Наука о данных (Data Science)» и представленная Программа способна реализовать их в объявленных требованиях.

5. Общая трудоёмкость дисциплины «Наука о данных (Data Science)» составляет 2 зачётные единицы (72 часа/из них практическая подготовка 2 ч.).

6. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Наука о данных (Data Science)» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.04.02 «Информационные системы и технологии» и возможность дублирования в содержании отсутствует.

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

8. Программа дисциплины «Наука о данных (Data Science)» предполагает занятия в интерактивной форме.

9. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.04.02 «Информационные системы и технологии».

10. 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 с оценкой, что соответствует статусу дисциплины, как дисциплины по выбору части, формируемой участниками образовательных отношений учебного цикла – Б1.В. ФГОС ВО направления 09.04.02 «Информационные системы и технологии».

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 09.04.02 «Информационные системы и технологии».

12. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Наука о данных (Data Science)» 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 «Наука о данных (Data Science)».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Наука о данных (Data Science)» ОПОП ВО по направлению 09.04.02 «Информационные системы и технологии», направленность **«Науки о данных (Data Science)»** (квалификация выпускника – магистр), 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)