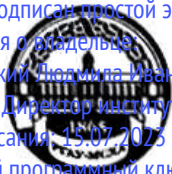


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Хоружий Людмила Ивановна
Должность: Директор института экономики и управления АПК
Дата подписания: 15.07.2022 19:17:36
Уникальный программный ключ:
1e90b132a5b04acc67585168b015ddd42cb1e6a9



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 управления АПК
Л. Хоружий
“ 30 ” *август* 2022 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.01 Основы науки о данных (Data Science)

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.02 «Информационные системы и технологии»
Направленность: Большие данные и машинное обучение (Machine Learning & Big Data);
Компьютерные науки и интеллектуальный анализ данных
(Computer Science & Data Mining)

Курс 1
Семестр 1

Форма обучения очная
Год начала подготовки 2022

Москва, 2022

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

(ФИО, ученая степень, ученое звание)

Ульянкин А.Е., ассистент


(ФИО, ученая степень, ученое звание)


«26» 08 2022 г.

«26» 08 2022 г.

Рецензент: Коломеева Е.С., к.э.н., доцент

(ФИО, ученая степень, ученое звание)


(подпись)
«26» 08 2022 г.

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

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 2022 г.

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 2022 г.

И.о. заведующего выпускающей кафедрой статистики и кибернетики

Уколова А.В., к.э.н., доцент

(ФИО, ученая степень, ученое звание)


(подпись)
«26» 08 2022 г.

/Заведующий отделом комплектования ЦНБ


Ермилова А.В.
(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	6
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ	6
ПО СЕМЕСТРАМ	6
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	9
4.3 ПРАКТИЧЕСКИЕ ЗАДАНИЯ.....	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	18
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	18
7.1 Основная литература	18
7.2 Дополнительная литература.....	19
8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	18
9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ	20
10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ.....	20
11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	21
12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ.....	21

Аннотация

рабочей программы по дисциплине Б1.В.01 «Основы науки о данных (Data Science)» для подготовки бакалавров по направлению 09.03.02 «Информационные системы и технологии» по направленности «Большие данные и машинное обучение (Machine Learning & Big Data)», «Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)»

Цель освоения дисциплины: Целью дисциплины «Основы науки о данных (Data Science)» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области поиска, критического анализа и синтеза информации с использованием языка программирования Python для решения поставленных задач.

Место дисциплины в учебном плане: дисциплина включена в часть, формируемую участниками образовательных отношений 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-1 (УК-1.1; УК-1.2; УК-1.3); ПКос-8 (ПКос-8.1; ПКос-8.2; ПКос-8.3).

Краткое содержание дисциплины:

Основные понятия и термины науки о данных. Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи). Математический аппарат. Введение в модуль math, NumPy. Основы работы с Pandas. Особенности фильтрации и обращения к данным. Модуль random. Загрузка данных в формате xlsx, json, txt. Ввод данных с клавиатуры. Логические выражения. Условный оператор. Инструкция if. Цикл While и For. Функции в программировании. Параметры и аргументы функций. Локальные и глобальные переменные. Процедуры. Разведывательный анализ данных

Ранжированный ряд распределения. Огиба распределения. Вариационный ряд распределения для дискретного признака. Полигон распределения. Интервальный вариационный ряд распределения. Гистограмма распределения. Показатели центральной тенденции. Средняя арифметическая простая и взвешенная. Структурные средние. Показатели вариации. Расчет показателей вариации с использованием встроенных функций. Написание собственных функций для расчета показателей вариации. Показатели асимметрии и эксцесса. Библиотека pandas.

Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования. Математические библиотеки: SciPy. Визуализация данных на Python.

Понятие статистической гипотезы. Основные этапы проверки, статистической гипотезы. Нулевая и рабочая гипотезы. Уровень значимости. Проверка гипотез относительно распределения численностей. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия.

Введение в машинное обучение. Задачи классификации и линейные модели. Обучение с учителем и без учителя. Проблема переобучения. Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг. Решение задач кластеризации. Меры расстояния. Обзор алгоритмов кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связанных компонент). Эффективные библиотеки работы с данными на Python: numpy, pandas. Машинное обучение на Python: sk-learn.

Общая трудоемкость дисциплины составляет: 2 зачетные единицы (72 часа).

Промежуточный контроль: зачет.

1. Цель освоения дисциплины

Целью дисциплины «Основы науки о данных (Data Science)» является освоение студентами теоретических и практических знаний и приобретение умений и навыков в области поиска, критического анализа и синтеза информации с использованием языка программирования Python для решения поставленных задач.

2. Место дисциплины в учебном процессе

Дисциплина «Основы науки о данных (Data Science)» включена в часть, формируемую участниками образовательных отношений учебного плана. Дисциплина «Основы науки о данных (Data Science)» реализуется в соответствии с требованиями ФГОС ВО, ОПОП ВО и Учебного плана по направлению 09.03.02 «Информационные системы и технологии».

Дисциплина «Основы науки о данных (Data Science)» является основополагающей для изучения следующих дисциплин: «Большие данные», «Математическая статистика», «Методы искусственного интеллекта», «Технологии обработки информации», «Многомерные статистические методы», «Интеллектуальный анализ данных и статистика на иностранном языке», «Интеллектуальный анализ данных и статистика».

Рабочая программа дисциплины Основы науки 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

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 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	УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Знать: методики поиска, сбора и обработки информации; актуальные российские и зарубежные источники информации в сфере профессиональной деятельности; метод системного анализа	методики поиска, сбора и обработки информации в сфере профессиональной деятельности		
			УК-1.2 Уметь: применять методики поиска, сбора и обработки информации; осуществлять критический анализ и синтез информации, полученной из разных источников; применять системный подход для решения поставленных задач		применять методики поиска, сбора и обработки информации в сфере профессиональной деятельности, проводить критический анализ информации	
			УК-1.3 Владеть: методами поиска, сбора и обработки, критического анализа и синтеза информации; методикой системного подхода для решения поставленных задач			методами поиска, сбора и обработки, критического анализа информации в сфере профессиональной деятельности
2.	ПКос-8	Способность проводить анализ данных с использованием информационных технологий в области сельского хозяйства	ПКос-8.1 Знать: основы технологии производства продукции сельского хозяйства; теорию и методологию дисциплин экономического профиля (экономика, бухгалтер-	информационные технологии анализа данных; источники информации для профессиональной деятельности		

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
		ства, экономики, бухгалтерского учета, статистики, финансов и др.	ский учет, статистика, финансы и др.); информационные технологии анализа данных; источники информации для профессиональной деятельности			
			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		выбирать и применять с использованием языка Python методы анализа данных, устанавливать причинно-следственные связи между признаками, делать выводы	
			ПКос-8.3 Владеть: методологией и навыками проведения анализа данных с использованием информационных технологий в области экономики, в том числе экономики сельского хозяйства			методологией и навыками проведения анализа данных с использованием языка Python

Таблица 2

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость	
	час. всего/*	В т.ч. по семестрам № 1
Общая трудоёмкость дисциплины по учебному плану	72/4	72/4
1. Контактная работа:	34,25/4	34,25/4
Аудиторная работа	34,25/4	34,25/4
<i>практические занятия (ПЗ)</i>	34/4	34/4
<i>контактная работа на промежуточном контроле (КРА)</i>	0,25	0,25
2. Самостоятельная работа (СРС)	37,75	37,75
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)</i>	28,75	28,75
<i>Подготовка к зачету</i>	9	9
Вид промежуточного контроля:	Зачет	

4.2 Содержание дисциплины

Таблица 3

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа		Внеаудиторная работа СР
		ПЗ/*	ПКР/*	
Раздел 1. Введение в науку о данных. Основы работы в Python	30	14/2		16
Раздел 2. Основные статистики в Python	11,75	6		5,75
Раздел 3. Графические возможности Python	6	2		4
Раздел 4. Статистические методы в Python	12	6		6
Раздел 5. Машинное обучение в Python	12	6/2		6
Контактная работа на промежуточном контроле (КРА)	0,25		0,25	
Итого по дисциплине	72	34/4	0,25	37,75

Раздел 1. Введение в науку о данных. Основы работы в Python.

Основные понятия и термины науки о данных. Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи). Математический аппарат. Введение в модуль math, NumPy. Основы работы с Pandas. Особенности фильтрации и обращения к данным. Модуль random. Загрузка данных в формате xls, json, txt. Ввод данных с

клавиатуры. Логические выражения. Условный оператор. Конструкция if. Цикл While и For. Функции в программировании. Параметры и аргументы функций. Локальные и глобальные переменные. Процедуры. Разведывательный анализ данных

Раздел 2. Основные статистики в Python

Ранжированный ряд распределения. Огиба распределения. Вариационный ряд распределения для дискретного признака. Полигон распределения. Интервальный вариационный ряд распределения. Гистограмма распределения. Показатели центральной тенденции. Средняя арифметическая простая и взвешенная. Структурные средние. Показатели вариации. Расчет показателей вариации с использованием встроенных функций. Написание собственных функций для расчета показателей вариации. Показатели асимметрии и эксцесса. Библиотека pandas.

Раздел 3. Графические возможности Python

Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования. Математические библиотеки: SciPy. Визуализация данных на Python.

Раздел 4. Статистические методы в Python

Понятие статистической гипотезы. Основные этапы проверки, статистической гипотезы. Нулевая и рабочая гипотезы. Уровень значимости. Проверка гипотез относительно распределения численностей. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия.

Раздел 5. Машинное обучение в Python

Введение в машинное обучение. Задачи классификации и линейные модели. Обучение с учителем и без учителя. Проблема переобучения. Дерево решений. Введение в ансамбли моделей. Случайный лес. Эффективные библиотеки работы с данными на Python: numpy, pandas. Машинное обучение на Python: sk-learn.

4.3 Практические занятия

Таблица 4

Содержание практических занятий и контрольные мероприятия

№ п/п	№ раздела	№ и название практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов
1	Раздел 1. Введение в науку о данных. Основы работы в Python	Практическая работа № 1. Основные термины и понятия науки о данных	УК-1.1 ПКос-8.1	Вопросы дискуссии	2
		Практическая работа № 2. Типы данных Python и методы работы с ними	УК-1.2 ПКос-8.1	Защита работы	4
		Практическая работа № 3. Введение в основные матема-	УК-1.2 ПКос-8.1	Защита работы	2

№ п/п	№ раздела	№ и название практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во часов
		тические модули Python			
		Практическая работа № 4. Управляющие конструкции в Python	УК-1.2 ПКос-8.1	Защита работы	2
		Практическая работа № 5. Основы работы с Pandas	УК-1.2 ПКос-8.1 ПКос-8.2	Защита работы	2
		Контрольная работа № 1	УК-1.2 ПКос-8.2	письменная работа	2
2	Раздел 2. Основные статистики в Python	Практическая работа № 6. Построение рядов распределения в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.2		2
		Практическая работа № 7. Расчет основных статистик в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.2	Защита работы	4
3	Раздел 3. Графические возможности Python	Практическая работа № 8. Построение графиков в Python	УК-1.1 УК-1.2 ПКос-8.1 ПКос-8.2	Защита работы	2
4	Раздел 4. Статистические методы в Python	Практическая работа № 9. Проверка гипотез относительно распределения совокупности в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.2 ПКос-8.3	Защита работы	2
		Практическая работа № 10. Корреляционно-регрессионный анализ в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.2 ПКос-8.3	Защита работы	4
5	Раздел 5. Машинное обучение в Python	Практическая работа № 11. Построение дерева решений в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.1 ПКос-8.2 ПКос-8.3	Защита работы	2
		Практическая работа № 12. Алгоритм «случайный лес» в Python	УК-1.1 УК-1.2 УК-1.3 ПКос-8.1 ПКос-8.2 ПКос-8.3	Защита работы	2
		Контрольная работа № 2	УК-1.1 УК-1.2 УК-1.3 ПКос-8.1 ПКос-8.2 ПКос-8.3	письменная работа	2

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	№ раздела и темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Раздел 1. Введение в науку о данных. Основы работы в Python	Особенности фильтрации и обращения к данным. Загрузка данных в формате json. Ввод данных с клавиатуры. (УК-1.1, УК-1.2, ПКос-8.1)
2.	Раздел 2. Основные статистики в Python	Вариационный ряд распределения для дискретного признака. Полигон распределения. Показатели асимметрии и эксцесса. Библиотека pandas. (УК-1.1, УК-1.2, ПКос-8.1)
3	Раздел 3. Графические возможности Python	Математические библиотеки: SciPy. (УК-1.1, УК-1.2, ПКос-8.1, ПКос-8.2)
4	Раздел 4. Статистические методы в Python	Логистическая регрессия. (УК-1.1, УК-1.2, УК-1.3, ПКос-8.1, ПКос-8.2, ПКос-8.3)
5	Раздел 5. Машинное обучение в Python	Обучение с учителем и без учителя. Проблема переобучения. (УК-1.1, УК-1.2, УК-1.3, ПКос-8.1, ПКос-8.2, ПКос-8.3)

5. Образовательные технологии**Применение активных и интерактивных образовательных технологий**

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий
1.	Практическая работа № 3. Введение в основные математические модули Python	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***Контрольная работа № 1***Вариант 1**

1. Создайте матрицу A:

2 3 1

3 5 2

4 8 15

Добавьте к этой матрице столбец справа с номером Вашего варианта

Создайте матрицу B:

6 6 2

4 7 6

5 8 7

Найти: $|A * E + A^T * B^{-1}|$

Найти произведение выделенных элементов матриц.

2. Создайте последовательность x из 100 случайных чисел. Отберите из них 10 чисел (y). Перемешайте вектор y. Выбрать случайное число от 2 до 15 и умножить его на 5-ый элемент вектора y.
3. Создать словарь W: 1- A
2- B

3-С

4-D

Вывести только обозначения букв. Вывести на печать только цифры.

Добавить словарь Z: 3-С
4-D
5-E

Объединить 2 словаря обновив при этом W.

Вариант 2

1. Создайте матрицу A:

7 3 1

6 9 2

4 8 5

Добавьте к этой матрице столбец справа с номером Вашего варианта.

Создайте матрицу B:

1 6 2

4 5 6

8 8 7

Найти: $|E * B^{-1} + A^T * B|$

Найти произведение выделенных элементов матриц.

2. Создайте последовательность x из 100 случайных чисел. Перемешайте вектор x. Отберите из них 10 чисел (y). Выбрать случайное число от 1 до 12 и умножить его на 7-ый элемент вектора y.
3. Создать словарь W:

1- один

2-два

3-три

Вывести только слова. Вывести на печать только цифры.

Добавить словарь Z: 3-три
4-четыре

Объединить 2 словаря обновив при этом Z.

Контрольная работа № 2

Вариант 1

1. Создайте вектор v: 2,3,4,2,3, 2,3,4,1,1,1,5,6,6,6.

Добавьте в конец вектора элементы 5,7,8

Отсортировать вектор по возрастанию. Построить огиву ранжированного ряда. Удалить элемент, стоящий на 4 месте в отсортированной списке. На вторую позицию вставить значение «б». Создать вектор x, состоящий из 20 случайных чисел. Объединить векторы x и y. Проверить соответствует ли он нормальному закону распределения получившийся вектор по критерию Шапиро. Сделать вывод.

2. Создать функцию, которая на вход берет число и выводит все значения от 1 до этого числа. Пример: Вводим 5. Выход:

1

2

3

4

5

3. Загрузить данные по своему варианту из файла «Данные». Определить корреляцию между ВРП на душу населения (y) и инвестициями в основной капитал на душу (x) населения. Построить уравнение регрессии y от x. Рассчитать по показателю x среднюю, медиану, дисперсию и коэффициент вариации. Сделать вывод об однородности совокупности.

4. Загрузить данные `student-por1.csv` в Python. Для переменной **G3** повести классификацию методами:

- дерево решений;

-случайный лес.

Вариант 2

1. Создайте вектор v : 1,3,4,1,3, 7,3,4,1,7,8,5,6,33,6.

Добавьте в конец вектора элементы 5,7,8. Отсортировать вектор по возрастанию. Построить огиу ранжированного ряда. Удалить максимальный элемент из ряда. На вторую позицию вставить значение «3». Создать вектор x , состоящий из 30 случайных чисел. Объединить векторы x и y . Проверить соответствует ли он нормальному закону распределения получившийся вектор по критерию Пирсона. Сделать вывод.

2. Создать функцию, которая на вход берет 2 числа x и y и выводит все значения от x до y . Пример: Вводим 2, 5. Выход:

2
3
4
5

Предусмотреть условие, что x должен быть больше y .

3. Загрузить данные по своему варианту из файла «Данные». Определить корреляцию между ВРП на душу населения (y) и инвестициями в основной капитал на душу (x) населения. Построить уравнение регрессии y от x . Рассчитать по показателю x среднюю, медиану, дисперсию и коэффициент вариации. Сделать вывод об однородности совокупности.

4. Загрузить данные `student-por2.csv` в Python. Для переменной **G3** повести классификацию методами:

- дерево решений;
- случайный лес.

Вопросы к зачету

1. Дайте определение понятия «Наука о данных (Data Science)».
2. Сущность понятий: данные, информация и показатель.
3. Что такое статистика? Отличия статистики от науки о данных.
4. В чем суть машинного обучения?
5. Искусственный интеллект: понятия и практическое применение.
6. Основные описательные статистики. Их интерпретация.
7. Статистические гипотезы: сущность, примеры.
8. Особенности применения корреляционно-регрессионного анализа.
9. Сущность интеллектуального анализа данных.
10. Отличия между глубоким и машинным обучением.
11. Сущность профессии Data Scientist.
12. Основные возможности языка программирования Python в науке о данных.
13. Язык программирования Python: типы данных, операции над числовыми типами данных.
14. Язык программирования Python: строки, списки, кортежи, словари.
15. Язык программирования Python: возможности библиотек `math`, `numpy` и `random`.
16. Язык программирования Python: статистические функции.
17. Графические возможности языка программирования Python.
18. Статистические гипотезы на языке программирования Python.
19. Корреляционно-регрессионный анализ в Python.

20. Методы машинного обучения на языке Python.
21. Классификация с помощью деревьев решений.
22. Классификация с помощью метода «Случайный лес».
23. Практическое применение методов науки о данных.

Пример работ

Практическая работа № 1. Основные термины и понятия науки о данных.

Вопросы для обсуждения:

1. Понятие «Наука о данных (Data Science)». Достоинства и недостатки различных определений понятия.
2. Сущность понятий: данные, информация и показатель.
3. Что такое статистика? Что Вы понимаете под математической статистикой?
4. Необходимость статистики в науке о данных.
5. Что такое машинное обучение? В чем суть машинного обучения?
6. Искусственный интеллект: понятия и практическое применение.

Практическая работа № 2. Типы данных Python и методы работы с ними.

Загрузите spyder. Создайте следующие типы данных: переменные, листы, словари, кортежи. Для вектора написать алгоритм расчета:

$$\sum v^2 + \frac{\max(b) - \min(b)}{\sum b} + \prod b$$

Построить в порядке убывания вектор b, найти произведение 5 и 2-го элементов в отсортированном списке.

Словарь W:

a-	A
б-	b
в-	v

Вывести только слова. Вывести на печать только русские буквы.

Добавить словарь Z:

с-	c
д-	d

Объединить 2 словаря обновив при этом Z.

Вопросы к защите:

1. В чем особенности работы со словарями.
2. Для чего могут использоваться словари.
3. Обращение к элементам векторов.

Практическая работа № 3. Введение в основные математические модули Python

Загрузите spyder. Напишите алгоритм расчета для любого x:

$$\frac{\sqrt[3]{(x+8)}}{\sin\left(\frac{x}{8}\right)} + \lg |x|$$

$$\sqrt{\frac{\cos x/2}{|x|}} + \ln x$$

Округлить в большую сторону.

Вопросы к защите:

1. Возможности модуля math.
2. Способы расчета логарифмов.
3. Тригонометрические функции.

Практическая работа № 4. Управляющие конструкции в Python

Написать программу для рисования горки. На вход подать 2 числа. 1 число – шаг горки. 2 число – высота горки.

Пример: 2 3

```
**  
*****  
*****
```

Вопросы к защите:

1. Особенности использования конструкции elif
2. Способы создания циклов
3. Выходы из циклов

Практическая работа № 5. Основы работы с Pandas

Загрузите spyder. Загрузите файл с исходными данными Вашего варианта в формате xlsx, csv и txt. При загрузке автоматически присвойте первой строке названия столбцам. Обратитесь только к 3 строке таблицы. Обратитесь к 2 и 54 столбцам таблицы. Рассчитайте сумму значений по каждому из столбцов.

Вопросы к защите:

1. Обращения к столбцам таблицы
2. Обращение к строкам таблицы
3. Добавление новых строк и столбцов.

Практическая работа № 6. Построение рядов распределения в Python

Загрузите spyder. Загрузите файл с исходными данными Вашего варианта. По показателю «Урожайность зерновых» построить ранжированный ряд распределения и его график (огиву). Построить интервальный ряд распределение и его график (гистограмму). Оформить отчет с выводами.

Вопросы к защите:

1. Особенности построения гистограмм в Python
2. Редактирование числа интервалов
3. Особенности построения ранжированного ряда в Python

Практическая работа № 7. Расчет основных статистик в Python

Загрузите spyder. По исходным данным предыдущей задачи рассчитать с использованием встроенных функций: среднее значение, медиану, дисперсию, коэффициент вариации, размах вариации и среднее квадратическое отклонение. Напишите собственные функции для расчета данных показателей. Сравните результаты.

Вопросы к защите:

1. Показатели центральной тенденции

2. Показатели вариации
3. Встроенные функции Python

Практическая работа № 8. Построение графиков в Python

Загрузите spyder. На сайте gks.ru найти данные для построения: диаграмм рассеяния, столбчатых диаграмм, круговых диаграмм, гистограммы и линейного графика. По каждому графику написать вывод. Обосновать выбор показателей. Оформить отчет с выводами.

Вопросы к защите:

1. Графические возможности Python
2. Редактирование графиков
3. Добавление нескольких графиков
4. Составные графики
5. Модуль matplotlib

Практическая работа № 9. Проверка гипотез относительно распределения совокупности в Python

По данным работы 6 проверить показатель на соответствие нормальному закону распределения по критериям: Шапиро, Пирсона, Колмогорова-Смирнова. Сравнить результаты. Оформить отчет с выводами.

Вопросы к защите:

1. Основные виды статистических проверок гипотез.
2. Критерий уровня значимости.
3. Общий метод построения доверительных интервалов и проверки статистических гипотез.
4. Смысл понятий «ошибка первого рода», «ошибка второго рода», «сила теста».

Практическая работа № 10. Корреляционно-регрессионный анализ в Python

Загрузите spyder. Загрузите файл с исходными данными Вашего варианта. Провести корреляционно-регрессионный анализ. Построить: парную модель регрессии и множественную модель. Сравнить результаты. Оформить отчет с выводами.

Вопросы к защите:

1. Оценка качества уравнения регрессии
2. Оценка значимости уравнения в целом
3. Оценка значимости параметров уравнения регрессии
4. Отбор факторов в уравнение регрессии
5. Предпосылки МНК

Практическая работа № 11. Построение дерева решений в Python

Загрузить исходные данные в Spyder. Зависимой переменной для построения дерева решений является итоговый результат поступления студента в вуз (поступил/не поступил). В качестве факторов используйте все остальные переменные. С помощью пакетов gpart и gpart.plot построить дерево решений двумя спо-

собами. Описать построенные деревья. Оценить качество классификации. Оформить отчет с выводами.

Вопросы к защите:

1. Виды деревьев решений
2. Интерпретация результатов
3. Сокращение числа «ветвей»
4. Оценка качества построенного дерева
5. Особенности бинарных деревьев решений

Практическая работа № 12. Алгоритм «случайный лес» в Python

По исходным данным предыдущей задачи построить модель классификации. Провести два варианта анализа: с обучением и без обучения. Сравнить результаты классификации. Сравнить качество классификации с предыдущими методами классификации. Оформить отчет с выводами.

Вопросы к защите:

1. Категориальные переменные
2. Особенности выбора зависимой переменной
3. Оценка качества модели
4. Интерпретация результатов
5. Применение как метод классификации

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

Текущий контроль знаний, умений и навыков проводится в форме 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%)

Критерии выставления оценок по системе:

- 0-59 % от максимального количества баллов – «незачтено»;
- 60 и более %– «зачтено».

7. Учебно-методическое и информационное обеспечение дисциплины

7.1 Основная литература

1. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 214 с. — (Высшее образование). — ISBN 978-5-534-15733-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/509562> (дата обращения: 24.09.2022).

2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469306> (дата обращения: 24.09.2022).
3. Гниденко, И. Г. Технологии и методы программирования : учебное пособие для вузов / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — Москва : Издательство Юрайт, 2022. — 235 с. — (Высшее образование). — ISBN 978-5-534-02816-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489920> (дата обращения: 24.09.2022).
4. Щерба, А. В. Программирование на Python. Первые шаги / А. В. Щерба. — Москва : Лаборатория знаний, 2022. — 250 с. — ISBN 978-5-93208-578-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/221678> (дата обращения: 13.09.2022). — Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — 5-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 538 с. — (Высшее образование). — ISBN 978-5-534-10004-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/495110> (дата обращения: 24.09.2022).
2. Кремер, Н. Ш. Эконометрика : учебник и практикум для вузов / Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 308 с. — (Высшее образование). — ISBN 978-5-534-08710-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/488678> (дата обращения: 24.09.2022).
3. 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: 24.09.2022). — Режим доступа: для авториз. пользователей.
4. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для среднего профессионального образования / Л. А. Станкевич. — Москва : Издательство Юрайт, 2022. — 397 с. — (Профессиональное образование). — ISBN 978-5-534-11659-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/495988> (дата обращения: 24.09.2022).

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

1. Язык программирования Python. URL: <https://www.python.org/>(открытый доступ)
2. Анаконда. URL: <https://www.anaconda.com/distribution/>(открытый доступ)
3. Официальный сайт Росстата. URL: <http://www.gks.ru/> (открытый доступ)
4. Официальный сайт Центрального Банка России. URL: <http://www.cbr.ru> (открытый доступ)

9. Перечень программного обеспечения

Таблица 9

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
4	Раздел 1. Введение в науку о данных. Основы работы в Python Раздел 2. Основные статистики в Python Раздел 3. Графические возможности Python Раздел 4. Статистические методы в Python Раздел 5. Машинное обучение в Python	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, учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (2й учебный корпус, 102 ауд.)</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 iP-4 541 3200 Mhz/1024 Mb/ 80 Gb / DVD-R 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 (2й учебный корпус, 302 ауд.)</i>	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.№).
Студенческое общежитие	Комнаты для самоподготовки
ЦНБ имени Н.И. Железнова	Читальный зал

11. Методические рекомендации студентам по освоению дисциплины

Приступая к изучению дисциплины «Основы науки 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.

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 «Основы науки о данных (Data Science)» предполагает изучение рекомендуемой учебно-методической литературы, подготовку к практическим занятиям, 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 зада-

ния занятия. 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, 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 освобождаться от сдачи зачета по дисциплине.

Программу разработал (и):

Харитонов А.Е., к.э.н., доцент

Ульянкин А.Е., ассистент



РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.01 «**Основы науки о данных (Data Science)**» ОПОП ВО по направлению 09.03.02 «**Информационные системы и технологии**», направленность «**Большие данные и машинное обучение (Machine Learning & Big Data)**», «**Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)**»
(квалификация выпускника – бакалавр)

Колосовой Еленой Сергеевной, 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 «**Основы науки о данных (Data Science)**» ОПОП ВО по направлению 09.03.02 «**Информационные системы и технологии**», направленность «**Большие данные и машинное обучение (Machine Learning & Big Data)**» и «**Компьютерные науки и интеллектуальный анализ данных (Computer Science & Data Mining)**» (уровень обучения - бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (разработчики – Харитоновна Анна Евгеньевна, кандидат экономических наук, доцент кафедры статистики и кибернетики и Ульянов Александр Евгеньевич, ассистент кафедры статистики и кибернетики).

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

1. Предъявленная рабочая программа дисциплины «**Основы науки о данных (Data Science)**» (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 части, формируемой участниками образовательных отношений учебного цикла учебного цикла — Б1.В.

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 «**Основы науки о данных (Data Science)**» закреплено 2 компетенции (6 индикаторов). Дисциплина «**Основы науки о данных (Data Science)**» 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 «**Основы науки о данных (Data Science)**» составляет 2 зачётные единицы (72 часа).

7. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «**Основы науки о данных (Data Science)**» 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 «**Основы науки о данных (Data Science)**» 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 зачета в 1 семестре, что соответствует статусу дисциплины, как дисциплины части, формируемой участниками образовательных отношений учебного цикла – Б1.В 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 – 4 источника (базовый учебник), дополнительной литературой – 4 наименования Интернет-ресурсы – 4 источника и соответствует требованиям ФГОС ВО направления 09.03.02 «Информационные системы и технологии».

14. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Основы науки о данных (Data Science)» 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 «**Основы науки о данных (Data Science)**».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Основы науки о данных (Data Science)» ОПОП ВО по направлению 09.03.02 «Информационные системы и технологии», направленность «**Большие данные и машинное обучение (Machine Learning & Big Data)**» и «**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 и Ульяновым А.Е. ассистентом 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

(подпись)



«24» августа 2022 г.