

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Хоружий, Татьяна Ивановна

Должность: Директор института экономики и управления АПК

Дата подписания: 01.08.2025 16:27:57

Уникальный идентификатор документа:

1e90b132d9b04dca67585160b015ddd2cb1e6a9

**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

И.И. Хоружий



«28» августа 2025 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Б1.В.14 Технологии работы с открытыми данными**

для подготовки бакалавров

ФГОС ВО

Направление: 09.03.03 Прикладная информатика

Направленность: Программные решения для бизнеса

Курс 3

Семестр 6

Форма обучения: очная

Год начала подготовки: 2025

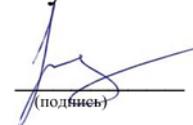
Москва, 2025

Разработчики: Калитвин В.А., канд. ф.-м. наук, доцент  
(ФИО, ученая степень, ученое звание)

  
(подпись)

«28» августа 2025 г.

Рецензент: Иноземцев А.И., канд. ф.-м. наук, доцент  
(ФИО, ученая степень, ученое звание)

  
(подпись)

«28» августа 2025 г.

Программа составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.03 Прикладная информатика, профессионального стандарта и учебного плана 2025 года начала подготовки.

Программа обсуждена на заседании кафедры статистики и кибернетики протокол № 11 от «26» августа 2025 г.

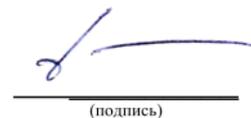
И.о. зав. кафедрой Уколова А.В., канд. экон. наук, доцент  
(ФИО, ученая степень, ученое звание)

  
(подпись)

«28» августа 2025 г.

**Согласовано:**

Председатель учебно-методической  
комиссии института экономики и управления АПК  
Гупалова Т.Н., канд. экон. наук, доцент  
(ФИО, ученая степень, ученое звание)

  
(подпись)

«28» августа 2025 г.

И.о. заведующего выпускающей кафедрой  
прикладной информатики  
Худякова Е.В., д.э.н., проф.  
(ФИО, ученая степень, ученое звание)

  
(подпись)

«28» августа 2025 г.

Заведующий отделом комплектования ЦНБ

  
(подпись)

## СОДЕРЖАНИЕ

<b>АННОТАЦИЯ .....</b>	<b>4</b>
<b>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>5</b>
<b>2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ .....</b>	<b>5</b>
<b>3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....</b>	<b>6</b>
<b>4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....</b>	<b>16</b>
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ .....	16
ПО СЕМЕСТРАМ .....	16
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	16
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ.....	17
<b>5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ .....</b>	<b>20</b>
<b>6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....</b>	<b>21</b>
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ .....	21
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ .....	51
<b>7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....</b>	<b>51</b>
7.1 ОСНОВНАЯ ЛИТЕРАТУРА .....	52
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	52
7.3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ, РЕКОМЕНДАЦИИ И ДРУГИЕ МАТЕРИАЛЫ К ЗАНЯТИЯМ.....	56
<b>8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕР-НЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ).....</b>	<b>56</b>
<b>9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ .....</b>	<b>56</b>
<b>10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ .....</b>	<b>57</b>
<b>11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ .....</b>	<b>61</b>
<b>12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ.....</b>	<b>62</b>

## АННОТАЦИЯ

### **рабочей программы учебной дисциплины Б1.В.14 Технологии работы с открытыми данными для подготовки бакалавров по направлению 09.03.03 Прикладная информатика, направленность «Программные решения для бизнеса»**

**Цель освоения дисциплины:** Целью дисциплины «Технологии работы с открытыми данными» является освоение студентами теоретических и практических знаний и приобретение умений и навыков работы с данными, их поиска, анализа, визуализации посредством использования языка программирования Python. По окончании изучения дисциплины студент должен знать: методы и приемы поиска и извлечения и анализа открытых данных; алгоритмы регрессии, классификации, кластерного анализа, нейронных сетей, области и способы их применения; синтаксис выбранного языка программирования, особенности программирования моделей машинного обучения на этом языке. По окончании изучения дисциплины студент должен уметь: использовать методы и приемы формализации и алгоритмизации поставленных задач; использовать инструменты визуализации данных; применять стандартные алгоритмы в соответствующих областях; применять выбранные языки программирования для написания программного кода; использовать выбранную среду программирования; применять инструментарий для создания и актуализации исходных текстов программ; выявлять ошибки в программном коде, интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов; применять методы и приемы отладки программного кода. Также по окончании изучения дисциплины студент должен владеть: навыками составления формализованных описаний решений поставленных задач в соответствии с требованиями технического задания; разработки алгоритмов решения поставленных задач в соответствии с требованиями технического задания или других принятых в организации нормативных документов; создания программного кода в соответствии с техническим заданием (готовыми спецификациями); оптимизации программного кода с использованием специализированных программных средств; анализа и проверки исходного программного кода; отладки программного кода на уровне программных модулей и межмодульных взаимодействий и взаимодействий с окружением.

**Место дисциплины в учебном плане:** дисциплина включена в часть, формируемую участниками образовательных отношений, учебного плана по направлению подготовки 09.03.03 Прикладная информатика.

**Требования к результатам освоения дисциплины:** результате освоения дисциплины формируются следующие компетенции (индикаторы): УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD- 4).1; ПК-6 (BD- 4).2; ПК-6 (BD- 4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.

**Краткое содержание дисциплины:**

Введение в данные: источники, типы данных, качество данных. Правовые аспекты открытых данных. Открытые данные в России и за рубежом. Методология работы с данными. Основные описательные статистики: средняя, мода, медиана. Проверка и подготовка данных. Расчет описательных статистик на Python. Типы данных. Форматы данных: Json, csv, xml. Различия и методы работы и сферы применения. Загрузка данных различных форматов в среду программирования. Библиотека Pandas. Работа с Data Frame. Визуализация данных. Виды графиков, построение диаграмм в Python. Методы анализа данных в Python. Классификация, регрессия, машинное обучение (Линейная регрессия, KNN, Дерево решений, K-Means, Ансамбли моделей, логистическая регрессия и др.). Подбор параметров моделей машинного обучения.

**Общая трудоемкость дисциплины: 108 / 3 (часы/зач. ед.)**

**Промежуточный контроль: зачет**

### **1. Цель освоения дисциплины**

Целью дисциплины «Технологии работы с открытыми данными» является освоение студентами теоретических и практических знаний и приобретение умений и навыков работы с данными, их поиска, анализа, визуализации посредством использования языка программирования Python, а также формирование знаний в области алгоритмов машинного обучения. По результатам изучения дисциплины студент должен уметь применять различные алгоритмы машинного обучения и решать возникающие перед ним практические задачи.

### **2. Место дисциплины в учебном процессе**

Дисциплина «Технологии работы с открытыми данными» включена в часть, формируемую участниками образовательных отношений, учебного плана. Дисциплина «Технологии работы с открытыми данными» реализуется в соответствии с требованиями ФГОС ВО, ОПОП ВО и Учебного плана по направлению 09.03.03 Прикладная информатика.

С учетом КРМ ИИ и ОП для направленности «Программные решения для бизнеса» в качестве профессиональных ролей в рамках компетентностно-ролевой модели в сфере ИИ выбраны следующие профессиональные роли:

- архитектор данных Data Architect (MLOps);
- менеджер проектов ИИ (IM PM).

Дисциплина «Технологии работы с открытыми данными» изучается на третьем курсе образовательного цикла.

Для успешного изучения дисциплины необходимы знания и умения по предшествующим дисциплинам:

Математический анализ, Дискретная математика, Математическая статистика, Исследование операций и методы оптимизации, Алгоритмизация и программирование, Базы данных, Информационные системы и технологии, Линейная алгебра, Веб-технологии.

Дисциплина «Технологии работы с открытыми данными» является основополагающей для изучения следующих дисциплин: «Управление информаци-

онными системами в АПК», «Системы поддержки принятия решений в АПК», «ВІ-системы в экономике АПК».

Особенностью дисциплины является рассмотрение алгоритмов машинного обучения средствами языка программирования Python и его открытых библиотек для извлечения, анализа, обработки, визуализации данных, а также библиотек для работы с моделями машинного обучения.

Рабочая программа дисциплины «Технологии работы с открытыми данными» 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 дисциплины направлено на формирование у обучающихся профессиональных компетенций (ПК), представленных в таблице 1.

Таблица 1

**Требования к результатам освоения учебной дисциплины «Технологии работы с открытыми данными»**

№ 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; Знает принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач	Знает принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач		
			УК-1.2; Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности.		Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности.	
			УК-1.3; Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений.			Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений.

2	УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)	УК-4.1; Знает принципы построения устного и письменного высказывания на государственном и иностранном языках; требования к деловой устной и письменной коммуникации.	Знает принципы построения устного и письменного высказывания на государственном и иностранном языках; требования к деловой устной и письменной коммуникации.		
3	ПК-4 (BD-2)	Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных (продвинутый уровень)	ПК-4 (BD-2).1; Определяет требования к наборам и качеству данных для решения задач машинного обучения. Уровень освоения индикатора: Ставит задачу разметки и оценивает качество работы разметчиков	методы определения требований к наборам и качеству данных для решения задач машинного обучения, способы постановки задач разметки и оценки качества работы разметчиков	определять требования к наборам и качеству данных для решения задач машинного обучения, ставить задачу разметки и оценивать качество работы разметчиков	методами определения требований к наборам и качеству данных для решения задач машинного обучения, способами постановки задач разметки и оценки качества работы разметчиков
			ПК-4 (BD-2).2; 1. Работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность Уровень освоения индикатора:	методы работы с данными, в том числе сбора данных из разрозненных источников, проверки данных на корректность; методами подбора инструментария разметки под условия задачи; способами организации краудсорсинг разметки	работать с данными, в том числе собирать данные из разрозненных источников, проверять данные на корректность, методы подбора инструментария разметки под условия задачи; способы организации краудсорсинг разметки	навыками работы с данными, в том числе сбора данных из разрозненных источников, проверки данных на корректность, методами подбора инструментария разметки под условия задачи; способами организации краудсорсинг разметки

			подбирает инструментарий разметки под условия задачи; организует краудсорсинг разметки			
4	ПК-6 (BD- 4)	Способен применять различные модели и (или) технологии обработки данных (продвинутый уровень)	ПК-6 (BD- 4).1; Осуществляет выбор технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями Продвинутый Способен организовывать распределенное хранилище и параллельную обработку на базе современных технологий (Hadoop, Spark) больших данных	технологии обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями; способы организации распределенных хранилищ и параллельной обработки на базе современных технологий (Hadoop, Spark) больших данных	выбирать технологии обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями; рганизацовывать распределенные хранилища и параллельную обработку на базе современных технологий (Hadoop, Spark) больших данных	навыками выбора технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями; способами организации распределенных хранилищ и параллельной обработки на базе современных технологий (Hadoop, Spark) больших данных
			ПК-6 (BD- 4).2; Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных технологий обработ-	методы разработки и отладки прикладных решений с элементами ИИ с применением различных технологий обработки данных	разрабатывать и отлаживать прикладные решения с элементами ИИ с применением различных технологий обработки данных	навыками разработки и отладки прикладных решений с элементами ИИ с применением различных технологий обработки данных

			ки данных Продвинутый Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных тех- нологий обработки дан- ных			
			ПК-6 (BD- 4).3; Тестирует, испы- тывает и оцени- вает качество решений с эле- ментами ИИ, ре- ализованных с использованием технологий обра- ботки данных Продвинутый Испытывает решения с элементами ИИ параллельной и поточковой обработки распределенных данных	методы тестирования, испы- тания и оценки качества ре- шений с элементами ИИ, реализованных с использо- ванием технологий обработ- ки данных	тестировать, испытывать и оце- нивать качество решений с эле- ментами ИИ, реализованных с использованием технологий об- работки данных	навыками тестирования, ис- пытания и оценки качества решений с элементами ИИ, реализованных с использова- нием технологий обработки данных
5	ПК-7 (BD-5)	Способен применять тех- нологии организации ин- фраструктуры БД (про- двинутый уровень)	ПК-7 (BD-5).1; Осуществляет выбор направле- ния вспомога- тельных техноло- гических реше-	методы осуществления вы- бора направления вспомога- тельных технологических решений для формирования единого стека работы с большими данными для ре-	осуществлять выбор направле- ния вспомогательных техноло- гических решений для формиро- вания единого стека работы с большими данными для реше- ния поставленной задачи	навыками осуществления выбора направления вспомога- тельных технологических решений для формирования единого стека работы с большими данными для ре-

			<p>ний для формирования единого стека работы с большими данными для решения поставленной задачи</p> <p>Продвинутый</p> <p>Выполняет отдельные функции в проектах по созданию инфраструктуры БД</p>	<p>шения поставленной задачи</p>		<p>шения поставленной задачи</p>
			<p>ПК-7 (BD-5).2;</p> <p>Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных технологий организации инфраструктуры БД</p> <p>Продвинутый</p> <p>Участвует в разработке решений с элементами ИИ с применением различных технологий организации инфраструктуры БД</p>	<p>методы разработки и отладки прикладных решений с элементами ИИ с применением различных технологий организации инфраструктуры БД</p>	<p>разрабатывать и отлаживать прикладные решения с элементами ИИ с применением различных технологий организации инфраструктуры БД</p>	<p>навыками разработки и отладки прикладных решений с элементами ИИ с применением различных технологий организации инфраструктуры БД</p>

6	ПК-11 (LC-2)	Способен проводить эксперименты на данных формулировать гипотезы исследования строить (обучать дообучать) модели машинного обучения с оценкой их качества и анализом ошибок обеспечивать воспроизводимость и масштабируемость исследований на данных (базовый уровень)	ПК-11 (LC-2).1; Проводит эксперименты с моделями ИИ, оценивает их качество (точность, производительность) Базовый Проводит эксперименты с моделями, выдвигает гипотезы	методы проведения экспериментов с моделями ИИ, оценки их качества (точность, производительность)	проводить эксперименты с моделями ИИ, оценивать их качество (точность, производительность)	навыками проведения экспериментов с моделями ИИ, оценки их качества (точность, производительность)
			ПК-11 (LC-2).2; Проводит эксперименты на данных и визуализации результатов с применением технологий анализа данных (статистического анализа), методов и алгоритмов машинного обучения Базовый Создает план экспериментов на данных	методы проведения экспериментов на данных и визуализации результатов с применением технологий анализа данных (статистического анализа), методов и алгоритмов машинного обучения	проводить эксперименты на данных и визуализировать результаты с применением технологий анализа данных (статистического анализа), методов и алгоритмов машинного обучения	навыками проведения экспериментов на данных и визуализации результатов с применением технологий анализа данных (статистического анализа), методов и алгоритмов машинного обучения
7	ПК-12 (LC-3)	Способен проектировать и поддерживать архитектуру систем искусственного интеллекта (продвинутый уровень)	ПК-12 (LC-3).1; Создает и развивает архитектуры системы ИИ на всех этапах жизненного цикла Продвинутый	методы создания и развития архитектуры системы ИИ на всех этапах жизненного цикла	создавать и развивать архитектуру системы ИИ на всех этапах жизненного цикла	навыками создания и развития архитектуры системы ИИ на всех этапах жизненного цикла

			Применяет различные принципы и паттерны при проектировании архитектуры систем ИИ			
8	ПК-16 (ML-1)	Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ (продвинутый уровень)	ПК-16 (ML-1).1; Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ (продвинутый уровень) Продвинутый Анализирует и сопоставляет задачу с современными трендами, выделяет специфику задачи в контексте последних достижений ИИ	способы применения знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ (продвинутый уровень)	применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ (продвинутый уровень)	навыками применения знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ (продвинутый уровень)
			ПК-16 (ML-1).2; 2. Определяет тенденции разви-	методы определения тенденций развития, оценки новизны и практической	определять тенденции развития, оценивать новизну и практическую значимость своих решений	навыками определения тенденций развития, оценки новизны и практической зна-

			<p>тия, оценивает новизну и практическую значимость своих решений с точки зрения современного искусственного интеллекта</p> <p>Продвинутый</p> <p>Объясняет причины появления концепции больших данных (БД), разницу определений.</p> <p>Выявляет различные категории проблем больших данных с примерами</p> <p>Анализирует динамику появления новых технологий, сопоставляет собственные решения с современными исследованиями и индустриальными стандартами</p>	<p>значимости своих решений с точки зрения современного искусственного интеллекта</p>	<p>с точки зрения современного искусственного интеллекта</p>	<p>чимости своих решений с точки зрения современного искусственного интеллекта</p>
--	--	--	---	---	--	--

			<p>ПК-16 (ML-1).3. 3. Оценивает конкурирующие решения и разработки с точки зрения трендов современного искусственного интеллекта</p> <p>Продвинутый Оценивает конкурентные решения с учётом современных трендов (например, использование современных архитектур, подходов к интерпретируемости, устойчивости, энергоэффективности), анализирует преимущества и ограничения</p>	<p>методы оценки конкурирующих решений и разработки с точки зрения трендов современного искусственного интеллекта</p>	<p>оценивать конкурирующие решения и разработки с точки зрения трендов современного искусственного интеллекта</p>	<p>навыками оценки конкурирующих решений и разработки с точки зрения трендов современного искусственного интеллекта</p>
--	--	--	--	---	---	---

## 4. Структура и содержание дисциплины

### 4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 3 зач.ед. (108 часов), их распределение по видам работ и семестрам представлено в таблице 2.

Таблица 2

#### Распределение трудоёмкости дисциплины по видам работ 6 семестре

Вид учебной работы	Трудоёмкость	
	час. всего/*	в т.ч. по семестрам
		№ 3
<b>Общая трудоёмкость</b> дисциплины по учебному плану	<b>108</b>	<b>108</b>
<b>1. Контактная работа:</b>	<b>50,25</b>	<b>50,25</b>
<b>Аудиторная работа</b>	<b>50</b>	<b>50</b>
<i>лекции (Л)</i>	<b>16</b>	16
<i>практические занятия (ПЗ)</i>	<b>34/4</b>	34/4
<i>контактная работа на промежуточном контроле (КРА)</i>	<b>0,25</b>	0,25
<b>2. Самостоятельная работа (СРС)</b>	<b>57,75</b>	<b>57,75</b>
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям)</i>	<b>57,75</b>	57,75
Вид промежуточного контроля:	Зачет	

\* в том числе практическая подготовка.(см учебный план)

### 4.2 Содержание дисциплины

Таблица 3

#### Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ всего/*	ПКР	
Раздел 1. Открытые данные. Правовые аспекты. Источники данных.	8	2	2	-	4
Раздел 2. Методы работы с данными. Описательные статистики.	18,75	2	8	-	8,75
Раздел 3. Типы данных. Загрузка данных. Работа с таблицами.	19	2	2	-	15
Раздел 4. Визуализация данных.	14	2	2	-	10
Раздел 5. Обработка данных.	48	8	20/4	-	20
Контактная работа на промежуточном контроле (КРА)	0,25	-	-	0,25	-
<b>Всего за 6 семестр</b>	<b>108</b>	<b>16</b>	<b>34</b>	<b>0,25</b>	<b>57,75</b>
<b>Итого по дисциплине</b>	<b>108</b>	<b>16</b>	<b>34</b>	<b>0,25</b>	<b>57,75</b>

## Раздел 1. Открытые данные. Правовые аспекты. Источники данных.

Основные понятия и термины открытых данных. Особенности работы с данными. Открытые данные и исследования. Источники данных, их виды. Достоверность источников данных. Качество данных. Правовое регулирование открытых данных. Открытые данные в России. Открытые данные за рубежом.

## Раздел 2. Методы работы с данными. Описательные статистики.

Основные статистики работы с открытыми данными. Ряды распределения. Среднее значение, мода, медиана. Показатели вариации. Расчет описательных статистик на Python.

## Раздел 3. Типы данных. Загрузка данных. Работа с таблицами.

Основные типы данных, форматы данных. Json, csv, xml и другие форматы. Способы работы с данными. Различия и применение форматов. Загрузка данных в Python. Работа с DataFrame. Работа с типом Series.

## Раздел 4. Визуализация данных.

Системы визуализации данных. Визуализация данных в Python. Библиотеки для визуализации данных. Анализ на основе визуального представления.

## Раздел 5. Обработка данных.

Введение в машинное обучение. Задачи классификации и линейные модели. Обучение с учителем и без учителя. Проблема переобучения. Дерево решений. Ансамбли моделей. Эффективные библиотеки работы с данными на Python. Машинное обучение с применением Python.

### 4.3 Практические занятия

Таблица 4

#### Содержание практических занятий и контрольные мероприятия

Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во Часов/ из них практическая подготовка
Раздел 1. Открытые данные. Правовые аспекты. Источники данных.	Лекция 1. «Основные термины и понятия открытых данных».	УК-1.1; УК-1.2; УК-1.3; УК-4.1;	-	2
	Практическая работа 1. «Основные термины и понятия открытых данных».	ПК-4 (BD-2).1; ПК-4 (BD-2).2;	Устный опрос, Тест	2
Раздел 2. Методы работы с данными. Описательные статистики.	Лекция 2. Методы работы с данными. Описательные статистики.	УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1;	-	2
	Практическая работа 2. «Ав-	ПК-4 (BD-2).2;	Защита	2

Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во Часов/ из них практическая подготовка
	томатизированные способы получения открытых данных»	ПК-6 (BD- 4).1; ПК-6 (BD- 4).2; ПК-6 (BD- 4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.	ра-боты, Тест	2
	Практическая работа 3. «Построение рядов распределения в Excel, Python».		Защита ра-боты, Тест	
	Практическая работа 4. «Описательные статистики: средняя, мода, медиана в Excel, Python»		Защита ра-боты, Тест	
	Практическая работа 5. «Показатели вариации в Excel, Python»		Защита ра-боты, Тест	
Раздел 3. Типы данных. Загрузка данных. Работа с таблицами.	Лекция 3. Типы данных. Загрузка данных. Работа с таблицами.	УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD- 4).1; ПК-6 (BD- 4).2; ПК-6 (BD- 4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.	-	2
	Практическая работа 6. «Загрузка данных в Python, работа с типом Series. работа с Data Frame»		Защита ра-боты, Тест	2
Раздел 4. Визуализация данных.	Лекция 4. Визуализация данных.	УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD- 4).1; ПК-6 (BD- 4).2; ПК-6 (BD- 4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.	-	2
	Практическая работа 7. «Библиотеки для визуализации данных в Python»		Защита ра-боты, Тест	2
Раздел 5. Обра-	Лекция 5. Линейная регрессия.	УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1;	Защита ра-	2
	Практическая работа 8. «Построение моделей корреля-			4/2

Название раздела, темы	№ и название лекций/ практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во Часов/ из них практическая подготовка
ботка данных.	ционно-регрессионного анализа»	ПК-4 (BD-2).2; ПК-6 (BD- 4).1;	боты, Тест, Кейс	
	Лекция 6. Классификация методом машинного обучения «Дерево решений». Ансамбли моделей.	ПК-6 (BD- 4).2; ПК-6 (BD- 4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2;		2
	Практическая работа 9. «Классификация методом машинного обучения «Дерево решений»»	ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2;	Защита ра- боты, Тест, Кейс	4/2
	Лекция 7. Метрические алгоритмы машинного обучения. Алгоритмы KNN и KMeans.	ПК-16 (ML-1).3.		2
	Практическая работа 10. «Другие методы машинного обучения. Алгоритм KNN.»		Защита ра- боты, Тест	2
	Практическая работа 11. «Другие методы машинного обучения. Кластерный анализ. Алгоритм KMeans.»		Защита ра- боты, Тест, Кейс	2
	Практическая работа 12. «Оптимизация моделей, подбор параметров»		Защита ра- боты, Тест	2
	Практическая работа 13. «Ансамбли моделей»		Защита ра- боты, Тест	2
	Лекция 8. «Нейронные сети»		-	2
	Практическая работа 14. «Нейронные сети»		Защита ра- боты, Тест, Кейс	4
	Контрольная работа		Контроль- ная работа	2

**Перечень вопросов для самостоятельного изучения дисциплины**

<b>№ п/п</b>	<b>Название раздела, темы</b>	<b>Перечень рассматриваемых вопросов для самостоятельного изучения</b>
1.	Раздел 1. Открытые данные. Правовые аспекты. Источники данных.	Открытые данные за рубежом. Правовое регулирование открытых данных (отечественный и зарубежный опыт) (УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ).
2.	Раздел 2. Методы работы с данными. Описательные статистики.	Обработка данных в Excel: построение рядов распределения, расчет основных статистик, расчет показателей вариации. (УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD-4).1; ПК-6 (BD-4).2; ПК-6 (BD-4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.).
3.	Раздел 3. Типы данных. Загрузка данных. Работа с таблицами.	Библиотеки для загрузки данных, работа с Data Frame. (УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD-4).1; ПК-6 (BD-4).2; ПК-6 (BD-4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.).
4.	Раздел 4. Визуализация данных.	Библиотеки для построения графиков и диаграмм (ПК- (УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD-4).1; ПК-6 (BD-4).2; ПК-6 (BD-4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.).
5.	Раздел 5. Обработка данных.	Методы машинного обучения: регрессия, классификация, кластерный анализ, нейронные сети и др. (УК-1.1; УК-1.2; УК-1.3; УК-4.1; ПК-4 (BD-2).1; ПК-4 (BD-2).2; ПК-6 (BD-4).1; ПК-6 (BD-4).2; ПК-6 (BD-4).3; ПК-7 (BD-5).1; ПК-7 (BD-5).2; ПК-11 (LC-2).1; ПК-11 (LC-2).2; ПК-12 (LC-3).1; ПК-16 (ML-1).1; ПК-16 (ML-1).2; ПК-16 (ML-1).3.).

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

<b>№ п/п</b>	<b>Тема и форма занятия</b>	<b>Наименование используемых активных и интерактивных образовательных технологий (форм обучения)</b>
1.	Раздел 1. Открытые данные. Правовые аспекты. Источники данных.	ПЗ Анализ конкретных ситуаций, тесты, мультимедийные презентации, видео и аудиоматериалы
2.	Раздел 2. Методы работы с данными. Описательные статистики.	ПЗ Анализ конкретных ситуаций, интерактивные задания и тесты, мультимедийные презентации, видео и аудиоматериалы
3.	Раздел 3. Типы данных. Загрузка данных. Работа с таблицами.	ПЗ Анализ конкретных ситуаций, интерактивные задания и тесты, мультимедийные презентации, видео и аудиоматериалы

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий (форм обучения)
4.	Раздел 4. Визуализация данных.	ПЗ Анализ конкретных ситуаций, интерактивные задания и тесты, мультимедийные презентации, видео и аудиоматериалы
5.	Раздел 5. Обработка данных.	ПЗ Анализ конкретных ситуаций, интерактивные задания и тесты, мультимедийные презентации, видео и аудиоматериалы

## **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**

#### **Пример контрольной работы**

##### **Вариант 1**

1. Загрузить данные из файла.
2. Отсортировать вектор по возрастанию. Построить огиву ранжированного ряда. Удалить элемент, стоящий на 4 месте в отсортированной списке. На вторую позицию вставить значение «6». 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.
3. Создать функцию, которая автоматически строила дискретный ряд распределения.
4. Создать функцию, которая автоматически рассчитывала бы среднюю, моду, медиану в дискретном ряду распределения.
5. Создать функцию, которая автоматически рассчитывала бы показатели вариации в дискретном ряду распределения.
6. Построить модель парной линейной регрессии между показателями  $Y$  и  $X$ , дать интерпретацию коэффициенту полной регрессии, сделать прогноз на 5 лет вперед.
7. Построить модель машинного обучения по методу «Дерево решений», оценить точность классификации разными метриками.

##### **Вариант2**

1. Загрузить данные из файла.
2. Отсортировать вектор по возрастанию. Построить огиву ранжированного ряда. Удалить элемент, стоящий на 8 месте в отсортированной списке. На вторую позицию вставить значение «321». Создать вектор  $x$ , состоящий из 15 случайных чисел. Объединить векторы  $x$  и  $y$ . Проверить соответствует ли он нормальному закону распределения получившийся вектор по критерию Шапиро. Сделать вывод.
3. Создать функцию, которая автоматически строила интервальный ряд распределения.

4. Создать функцию, которая автоматически рассчитывала бы среднюю, моду, медиану в интервальном ряду распределения.
5. Создать функцию, которая автоматически рассчитывала бы показатели вариации в интервальном ряду распределения.
6. Построить модель парной линейной регрессии между показателями  $Y$  и  $X$ , дать интерпретацию коэффициенту полной регрессии, сделать прогноз на 5 лет вперед.
7. Построить модель машинного обучения по методу «Случайный лес», оценить точность классификации разными метриками.

### **Вопросы к зачету**

1. Понятие открытых данных.
2. Источники данных и их виды.
3. Качество данных.
4. Достоверность источников данных.
5. Правовое регулирование данных.
6. Открытые данные в России.
7. Открытие данные за рубежом.
8. Основные статистики данных.
9. Ряды распределения и их виды.
10. Ранжированные ряд распределения, когда используется, алгоритм построения.
11. Дискретный ряд распределения, сфера применения, алгоритм построения.
12. Интервальный ряд распределения, применение и алгоритм построения.
13. Показатели центральной тенденции.
14. Среднее значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.
15. Модальное значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.
16. Медианное значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.
17. Показатели вариации.
18. Размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение. Интерпретация и алгоритм расчетов в дискретном ряду распределения.
19. Коэффициент осцилляции, относительно линейное отклонение, коэффициент вариации. Интерпретация и алгоритм расчетов в дискретном ряду распределения.
20. Размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение. Интерпретация и алгоритм расчетов в интервальном ряду распределения.
21. Коэффициент осцилляции, относительно линейное отклонение, коэффициент вариации. Интерпретация и алгоритм расчетов в интервальном ряду распределения.

## Пример практических работ

### Практическая работа № 1. «Основные термины и понятия открытых данных».

*Вопросы для обсуждения:*

1. Понятие открытых данных.
2. Источники данных и их виды.
3. Качество данных.
4. Достоверность источников данных.
5. Правовое регулирование данных.
6. Открытые данные в России.
7. Открытие данные за рубежом.

### Практическая работа № 2. «Автоматизированные способы получения открытых данных».

1. Разработать программу для автоматизированного получения данных с веб-страниц (парсер).

2. Разработать программу для получения данных с применением API.

*Вопросы к защите:*

1. Структура веб-документа.
2. API.

### Практическая работа 3. «Построение рядов распределения в Excel, Python».

1. На основе исходных данных, выбрать показатели, подходящие для построения ранжированного, дискретного и интервального рядов распределения.

2. Написать программу, позволяющую автоматизировать процесс построения ранжированного, дискретного и интервального рядов распределения на Python.

3. Проверить правильность построения в Excel.

*Вопросы к защите:*

1. Основные статистики данных.
2. Ряды распределения и их виды.

### Практическая работа 4. «Описательные статистики: средняя, мода, медиана в Excel, Python»

1. На основе результатов выполнения предыдущего задания, написать программу, позволяющую автоматизировать процесс расчетов показателей центральной тенденции для ранжированного, дискретного и интервального рядов распределения на Python.

2. Проверить правильность построения в Excel.

*Вопросы к защите:*

1. Показатели центральной тенденции.
2. Среднее значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.

3. Модальное значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.

4. Медианное значение. Расчет в ранжированном, дискретном и интервальном ряду распределения.

### **Практическая работа 5. «Показатели вариации в Excel, Python»**

1. На основе результатов выполнения предыдущих заданий, написать программу, позволяющую автоматизировать процесс расчетов показателей вариации для ранжированного, дискретного и интервального рядов распределения на Python.

2. Проверить правильность построения в Excel.

*Вопросы к защите:*

1. Показатели вариации.

2. Размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение. Интерпретация и алгоритм расчетов в дискретном ряду распределения.

3. Коэффициент осцилляции, относительно линейное отклонение, коэффициент вариации. Интерпретация и алгоритм расчетов в дискретном ряду распределения.

4. Размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение. Интерпретация и алгоритм расчетов в интервальном ряду распределения.

5. Коэффициент осцилляции, относительно линейное отклонение, коэффициент вариации. Интерпретация и алгоритм расчетов в интервальном ряду распределения.

### **Практическая работа 6. «Загрузка данных в Python, работа с типом Series. работа с Data Frame»**

1. Загрузите файл с исходными данными Вашего варианта в формате xlsx, csv или txt. При загрузке автоматически присвойте первой строке названия столбцам.

2. Обратитесь только к 3 строке таблицы.

3. Обратитесь к 2 и 54 значениям таблицы.

4. Вывести первые 7 значений.

5. Вывести последние 12 значений.

6. Вместо поменять местами четные значения.

7. Добавить на каждое нечетное значение, данные по соседнему варианту.

8. Рассчитать среднюю, моду, медиану, дать интерпретацию.

9. Проверить на соответствие нормальному закону распределения.

*Вопросы к защите:*

1. Способы загрузки данных в среду.

2. Библиотеки для работы с таблицами.

3. Обращения к столбцам таблицы.

4. Обращение к строкам таблицы.
5. Добавление новых строк.

### **Практическая работа 7. «Библиотеки для визуализации данных в Python»**

1. По данным практической работы 6 подстроить линейный график, точечную дигамму, столбчатую диаграмму, круговую диаграмму, гистограмму.
2. Изменить настройки отображения диаграмм.

*Вопросы к защите:*

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

### **Практическая работа 8. «Построение моделей корреляционно-регрессионного анализа»**

1. По данным предыдущих работ построить парные и множественные модели регрессии.
2. Оценить значимость моделей и параметров.
3. Отобрать качественные модели, дать интерпретацию коэффициенту полной регрессии.
4. Сделать прогноз.

*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 9. «Классификация методом машинного обучения «Дерево решений»»**

1. Загрузить исходные данные.
2. Провести обучение модели методом «Дерево решений» на полном массиве данных.
3. На основе обученной модели провести классификацию данных.
4. Оценить точность классификации различными метриками.

*Вопросы к защите:*

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

### **Практическая работа 10. «Другие методы машинного обучения. Алгоритм KNN.»**

1. Загрузить исходные данные.
2. Провести обучение модели с применением алгоритма «KNN» на полном массиве данных.

3. На основе обученной модели провести классификацию данных.
4. Оценить точность классификации различными метриками.

*Вопросы к защите:*

1. Алгоритм KNN.
2. Интерпретация результатов.
3. Подбор оптимального  $k$ .
4. Оценка качества работы алгоритма KNN.

### **Практическая работа № 11. «Другие методы машинного обучения. Кластерный анализ. Алгоритм KMeans»**

1. На основе данных предыдущих задач провести обучение моделей применением алгоритма «KMeans».
2. Оценить точность модели.

*Вопросы к защите:*

1. Кластерный анализ.
2. Метрики и метрические пространства.
3. Метод «локтя» подбора параметра  $k$ .
3. Оценка качества модели.
4. Интерпретация результатов.

### **Практическая работа 12. «Оптимизация моделей, подбор параметров»**

1. На основе данных предыдущих задач провести оптимизацию моделей «Дерево решений», путем подбора «лучшего» критерия разбиения и глубины дерева.
2. Сравнить результаты с задачами 12 и 13.

*Вопросы к защите:*

1. Оптимальная модель
2. Критерий Джини
3. Критерий Энтропии
4. Глубина дерева

### **Практическая работа 13. «Ансамбли моделей»**

1. На основе данных предыдущих задач построить модель «случайный лес».
2. Сравнить результаты с результатами предыдущих задач.

*Вопросы к защите:*

1. Ансамбли моделей: бэггинг, стэкинг, бустинг.
2. Кроссвалидация.

## **Практическая работа 14. «Нейронные сети»**

1. На основе данных предыдущих задач построить нейронную сеть для решения задачи регрессии.

2. На основе данных предыдущих задач построить нейронную сеть для решения задачи классификации.

2. Сравнить результаты с результатом предыдущих задач.

*Вопросы к защите:*

1. Структура простейшей нейронной сети. Персептрон Розенблатта.

2. Многослойные нейронные сети.

3. Обучение нейронной сети. Метод градиентного спуска. Метод обратного распространения ошибки.

## **Тесты**

### **Тема 1. Основные термины и понятия открытых данных.**

Вопрос 1.

Что такое открытые данные?

А) Данные, доступные только сотрудникам государственных учреждений

Б) Данные, которые в машиночитаемом формате без ограничений доступны для свободного использования, распространения и переиспользования

В) Данные, опубликованные исключительно на сайтах муниципальных администраций

Правильный ответ: Б

Вопрос 2.

Какой из следующих форматов данных считается машиночитаемым?

А) PDF-файл со сканированной таблицей

Б) Изображение в формате JPG с диаграммой

В) CSV-файл со структурированной информацией

Правильный ответ: В

Вопрос 3.

Что означает принцип "Открытое правительство" (Open Government)?

А) Публикация всех государственных документов без исключения в печатном виде

Б) Открытость, прозрачность и подотчётность государственных органов через публикацию данных и взаимодействие с гражданами

В) Бесплатный доступ граждан только к законодательным актам

Правильный ответ: Б

Вопрос 4.

Какие из перечисленных источников являются порталными платформами открытых данных?

А) Google Drive, Dropbox и OneDrive

Б) data.gov.ru, data.europa.eu и открытые государственные реестры

В) GitHub, GitLab и Bitbucket

Правильный ответ: Б

Вопрос 5.

Что предполагает лицензирование открытых данных?

А) Полный запрет на использование данных третьими лицами

Б) Установление условий использования, переиспользования и распространения данных при сохранении авторства и соблюдении требований лицензии

В) Платёж за каждое использование набора данных

Правильный ответ: Б

## **Тема 2. Автоматизированные способы получения открытых данных.**

Вопрос 1

Что такое API (Application Programming Interface) в контексте работы с открытыми данными?

А) Программный интерфейс для автоматизированного доступа к данным с соблюдением установленных правил и форматов запросов

Б) Физический документ с инструкциями по использованию наборов данных

В) Лицензионное соглашение между производителем и потребителем данных

Правильный ответ: А

Вопрос 2

Какой из следующих форматов данных наиболее удобен для автоматизированной обработки и парсинга?

А) PDF-документ с неструктурированным текстом

Б) JSON или XML с иерархической структурой данных

В) Изображение в формате PNG

Правильный ответ: Б

Вопрос 3

Что означает веб-скрейпинг (web scraping) при работе с открытыми данными?

А) Автоматизированное извлечение структурированных данных с веб-сайтов с помощью программных средств

Б) Ручное переписывание информации с веб-страниц в электронную таблицу

В) Удаление данных с сервера портала открытых данных

Правильный ответ: А

Вопрос 4

Какие инструменты и библиотеки часто используются для автоматизированного получения открытых данных в Python?

- А) requests, BeautifulSoup, pandas и selenium
- Б) Microsoft Word, Excel и PowerPoint
- В) Adobe Photoshop, Illustrator и InDesign

Правильный ответ: А

Вопрос 5

Что такое REST API в контексте доступа к открытым данным?

- А) Архаичный метод передачи данных через телефонные линии
- Б) Архитектурный стиль, позволяющий получать данные через стандартные HTTP-запросы (GET, POST) к определённым URL-адресам
- В) Закрытая система доступа, требующая специального разрешения администратора

Правильный ответ: Б

### **Тема 3. «Построение рядов распределения в Python».**

Вопрос 1

Что такое ряд распределения?

- А) Графическое изображение зависимости между двумя переменными
- Б) Упорядоченное отображение единиц совокупности в порядке возрастания или убывания значения признака
- В) Таблица с результатами линейной регрессии

Правильный ответ: Б

Вопрос 2

Какая функция pandas используется для подсчёта частот значений в серии данных?

- А) count() — подсчёт количества ненулевых элементов
- Б) value\_counts() — подсчёт уникальных значений и их частоты
- В) describe() — описание статистических характеристик

Правильный ответ: Б

Вопрос 3

Какой метод pandas применяется для группировки данных в интервалы (например, по возрасту)?

- А) groupby() — группировка по категориальному признаку
- Б) cut() — разбиение непрерывного признака на интервалы с равной шириной
- В) merge() — объединение двух датафреймов

Правильный ответ: Б

Вопрос 4

Какой график визуализирует частотное распределение непрерывной переменной?

- А) График рассеивания (scatter plot)
- Б) Гистограмма (histogram) с полигоном распределения
- В) Линейный график (line plot)

Правильный ответ: Б

Вопрос 5

Что означает "относительная частота" в ряду распределения?

- А) Разница между максимальным и минимальным значением признака
- Б) Отношение частоты значения к общему числу наблюдений, выраженное в процентах или долях
- В) Стандартное отклонение от среднего значения

Правильный ответ: Б

#### **Тема 4. «Описательные статистики: средняя, мода, медиана в Python».**

Вопрос 1

Что называется средней (mean) в описательной статистике?

- А) Значение, которое встречается наиболее часто в наборе данных
- Б) Значение в середине упорядоченного ряда данных
- В) Сумма всех значений, делённая на количество наблюдений

Правильный ответ: В

Вопрос 2

Какой показатель центральной тенденции наиболее устойчив к влиянию выбросов?

- А) Средняя (mean)
- Б) Медиана (median)
- В) Мода (mode)

Правильный ответ: Б

Вопрос 3

Какой метод Python используется для вычисления среднего значения списка чисел через встроенный модуль statistics?

- А) statistics.avg() — вычисление среднего
- Б) statistics.mean() — вычисление средней
- В) statistics.average() — вычисление среднего значения

Правильный ответ: Б

Вопрос 4

Что такое мода (mode) в наборе данных?

- А) Наиболее часто встречающееся значение в выборке
- Б) Среднее арифметическое всех значений
- В) Разность между максимальным и минимальным значением

Правильный ответ: А

Вопрос 5

Какая функция pandas используется для одновременного расчёта средней, медианы, стандартного отклонения и других описательных статистик?

- А) Series.stats() — получение статистики серии
- Б) Series.describe() — описание основных характеристик данных
- В) Series.summary() — сводка по серии

Правильный ответ: Б

### **Тема 5. «Показатели вариации в Excel, Python».**

Вопрос 1

Что такое размах вариации в наборе данных?

- А) Сумма квадратов отклонений от среднего значения
- Б) Разница между максимальным и минимальным значением в выборке
- В) Отношение стандартного отклонения к средней величине

Правильный ответ: Б

Вопрос 2

Какой показатель вариации характеризует среднее квадратическое отклонение значений от среднего?

- А) Дисперсия (variance)
- Б) Коэффициент вариации (coefficient of variation)
- В) Межквартильный размах (IQR)

Правильный ответ: А

Вопрос 3

Какая функция Excel используется для расчёта стандартного отклонения по выборке?

- А) STDEV() или STDEV.S() — стандартное отклонение выборки
- Б) VAR() или VAR.S() — дисперсия выборки
- В) RANGE() — размах вариации

Правильный ответ: А

Вопрос 4

Что означает высокий коэффициент вариации в распределении данных?

- А) Данные тесно сгруппированы вокруг среднего значения
- Б) Данные сильно рассеяны относительно среднего значения
- В) Распределение абсолютно симметрично

Правильный ответ: Б

Вопрос 5

Какой метод Python используется для расчёта стандартного отклонения через библиотеку NumPy?

- А) np.std() — стандартное отклонение массива
- Б) np.variance() — дисперсия массива
- В) np.deviation() — отклонение от среднего

Правильный ответ: А

## **Тема 6. «Загрузка данных в Python, работа с типом Series. работа с Data Frame».**

Вопрос 1

Какая функция pandas используется для загрузки данных из CSV-файла?

- А) pd.load\_csv() — загрузка CSV-файла
- Б) pd.read\_csv() — чтение данных из CSV-файла

В) `pd.import_csv()` — импорт CSV-данных

Правильный ответ: Б

Вопрос 2

Что такое Series в pandas?

А) Одномерный массив с метками индекса и значениями одного типа данных

Б) Двумерная таблица с несколькими столбцами и строками

В) Функция для суммирования числовых данных

Правильный ответ: А

Вопрос 3

Какой метод используется для выбора одного столбца из DataFrame по названию?

А) `df['column_name']` — выбор столбца по названию

Б) `df.get_column('column_name')` — получение столбца

В) `df.select('column_name')` — выбор колонки

Правильный ответ: А

Вопрос 4

Что возвращает метод `df.info()` в DataFrame?

А) Список всех значений в таблице

Б) Информацию о структуре: типы данных, количество непустых значений, использование памяти

В) Первые 5 строк датафрейма

Правильный ответ: Б

Вопрос 5

Как получить размер (количество строк и столбцов) DataFrame?

А) `df.length` — длина датафрейма

Б) `df.size` — общее количество элементов

В) `df.shape` — кортеж (строки, столбцы)

Правильный ответ: В

### Вопрос 1

Какая библиотека Python является базовой для построения статических графиков и диаграмм?

- А) `matplotlib` — библиотека для построения двумерных графиков
- Б) `plotly` — библиотека для интерактивных визуализаций
- В) `seaborn` — библиотека только для работы со статистическими моделями

Правильный ответ: А

### Вопрос 2

Что такое `seaborn` и как она связана с `matplotlib`?

- А) Независимая библиотека, которая заменяет `matplotlib`
- Б) Высокоуровневая библиотека, построенная на основе `matplotlib`, с удобным API для статистической визуализации
- В) Расширение только для трёхмерной графики

Правильный ответ: Б

### Вопрос 3

Какой метод используется для построения гистограммы в `matplotlib`?

- А) `plt.histogram()` — построение гистограммы
- Б) `plt.hist()` — построение гистограммы
- В) `plt.plot_histogram()` — функция для гистограмм

Правильный ответ: Б

### Вопрос 4

Какая библиотека визуализации обеспечивает интерактивные графики с возможностью масштабирования и наведения указателя?

- А) `matplotlib` — только статические графики
- Б) `seaborn` — расширение для улучшенной статистики
- В) `plotly` — интерактивные и динамические визуализации

Правильный ответ: В

### Вопрос 5

Что делает функция `plt.show()` в `matplotlib`?

- А) Сохраняет график в файл PNG
- Б) Отображает построенный график в окне или в Jupyter Notebook
- В) Очищает все предыдущие графики из памяти

Правильный ответ: Б

## **Тема 8. «Построение моделей корреляционно-регрессионного анализа»**

Вопрос 1

Что такое коэффициент корреляции и какие значения он может принимать?

- А) Показатель, характеризующий силу и направление связи между двумя переменными, принимает значения от -1 до +1
- Б) Параметр регрессии, который всегда положительный и больше нуля
- В) Величина, показывающая влияние одной переменной на другую в причинно-следственном отношении

Правильный ответ: А

Вопрос 2

Какая из следующих корреляций указывает на сильную положительную связь между переменными?

- А)  $r = -0.85$  — сильная отрицательная связь
- Б)  $r = 0.92$  — сильная положительная связь
- В)  $r = 0.15$  — слабая положительная связь

Правильный ответ: Б

Вопрос 3

Что показывает коэффициент детерминации ( $R^2$ ) в регрессионной модели?

- А) Скорость изменения зависимой переменной
- Б) Долю дисперсии зависимой переменной, объяснённую независимой переменной
- В) Ошибку предсказания модели

Правильный ответ: Б

Вопрос 4

Какое из утверждений верно отражает связь между корреляцией и причинностью?

- А) Высокая корреляция всегда означает прямую причинно-следственную связь
- Б) Корреляция не означает причинность; высокая корреляция может быть результатом совпадения, конфаундинга или третьей переменной

В) Низкая корреляция указывает на наличие причинной связи

Правильный ответ: Б

Вопрос 5

Какой метод используется для построения модели простой линейной регрессии в Python (scikit-learn)?

- А) LinearRegression() из модуля sklearn.linear\_model для построения модели  $y = a + b \cdot x$
- Б) Regression() из встроенного модуля regression
- В) fit\_line() из библиотеки numpy

Правильный ответ: А

### **Тема 9. «Классификация методом машинного обучения «Дерево решений»»**

Вопрос 1

Что такое дерево решений в машинном обучении?

- А) Граф, состоящий из корневого узла, промежуточных узлов и листьев, которые используются для задач классификации и регрессии путём последовательного разбиения данных
- Б) Линейная модель, которая находит оптимальную прямую для разделения классов
- В) Тип нейронной сети с несколькими слоями

Правильный ответ: А

Вопрос 2

Какой критерий разбиения чаще всего используется в деревьях решений для задач классификации?

- А) Средняя функция потерь (mean squared error)
- Б) Коэффициент корреляции между признаками
- В) Индекс Джини (Gini impurity) или энтропия (entropy)

Правильный ответ: В

Вопрос 3

Что такое переобучение (overfitting) в контексте деревьев решений?

- А) Ситуация, когда дерево слишком простое и не может обучиться на данных
- Б) Слишком глубокое дерево, которое запоминает особенности обучающих данных и плохо обобщается на новые данные

В) Равномерное распределение данных по всем листьям дерева

Правильный ответ: Б

Вопрос 4

Какой параметр дерева решений контролирует максимальную глубину дерева в scikit-learn?

А) `max_features` — максимальное количество признаков

Б) `max_depth` — максимальная глубина дерева

В) `min_samples_split` — минимальное количество примеров для разбиения узла

Правильный ответ: Б

Вопрос 5

Какое преимущество имеет дерево решений перед другими методами классификации?

А) Требуется нормализация признаков перед обучением

Б) Легко интерпретируемо и может быть визуализировано; не требует нормализации данных

В) Гарантирует оптимальное разбиение данных

Правильный ответ: Б

## **Тема 10. «Другие методы машинного обучения. Алгоритм KNN»**

Вопрос 1

Что такое алгоритм KNN (k-ближайших соседей)?

А) Линейный алгоритм классификации, основанный на расчёте коэффициентов регрессии

Б) Алгоритм, классифицирующий объект на основе k ближайших к нему объектов в пространстве признаков

В) Метод кластеризации, который разбивает данные на иерархические группы

Правильный ответ: Б

Вопрос 2

Какой параметр является ключевым в алгоритме KNN?

А) `k` — количество ближайших соседей, которые используются для классификации

Б) `alpha` — коэффициент обучения (`learning rate`)

В) `epochs` — количество итераций обучения

Правильный ответ: А

Вопрос 3

Какие метрики расстояния часто используются в KNN для определения "близости" объектов?

- А) Евклидово расстояние (Euclidean) и манхэттенское расстояние (Manhattan)
- Б) Только дисперсия признаков
- В) Корреляция между векторами признаков

Правильный ответ: А

Вопрос 4

Почему перед использованием алгоритма KNN желательно нормализовать (масштабировать) признаки?

- А) Нормализация не влияет на результаты классификации
- Б) Для уменьшения влияния признаков с большими значениями на расчёт расстояния
- В) Чтобы увеличить количество признаков в модели

Правильный ответ: Б

Вопрос 5

Какой основной недостаток алгоритма KNN?

- А) Он требует нормализации данных перед обучением
- Б) Медленное предсказание, так как нужно рассчитывать расстояние до всех обучающих объектов; высокая чувствительность к шуму в данных
- В) Не может работать с категориальными признаками

Правильный ответ: Б

### **Тема 11. «Другие методы машинного обучения. Кластерный анализ. Алгоритм KMeans».**

Вопрос 1

Что такое кластерный анализ?

- А) Метод предсказания значений целевой переменной на основе признаков объекта
- Б) Задача разбиения множества объектов на группы (кластеры) на основе их сходства без предварительной информации о классах
- В) Проверка статистической значимости связи между переменными

Правильный ответ: Б

Вопрос 2

Как работает алгоритм KMeans?

- А) Итеративно перемещает центроиды кластеров и переназначает объекты к ближайшему центроиду до сходимости
- Б) Строит дерево иерархических разбиений данных снизу вверх
- В) Использует вероятностные модели для определения кластеров

Правильный ответ: А

Вопрос 3

Что означает параметр  $k$  в алгоритме KMeans?

- А) Количество признаков в датасете
- Б) Количество итераций обучения
- В) Количество кластеров, на которые необходимо разбить данные

Правильный ответ: В

Вопрос 4

Какой метод используется для определения оптимального числа кластеров в KMeans?

- А) Метод локтя (elbow method) — анализ зависимости инерции от числа кластеров
- Б) Только перебор всех возможных значений  $k$
- В) Используется произвольное число кластеров, выбранное пользователем

Правильный ответ: А

Вопрос 5

Почему результаты KMeans могут различаться при разных запусках алгоритма?

- А) KMeans имеет детерминированный результат, который всегда одинаков
- Б) Случайная инициализация центроидов может привести к разным локальным оптимумам
- В) Порядок входных данных влияет на результаты кластеризации

Правильный ответ: Б

## Тема 12. «Оптимизация моделей, подбор параметров»

Вопрос 1

Какое отличие между параметрами модели и гиперпараметрами?

- А) Параметры и гиперпараметры — это одно и то же
- Б) Параметры обучаются моделью на данных, а гиперпараметры устанавливаются пользователем перед обучением
- В) Гиперпараметры — это только для нейронных сетей

Правильный ответ: Б

Вопрос 2

Что такое переобучение (overfitting) модели?

- А) Модель хорошо обучена и имеет высокую точность на всех данных
- Б) Модель запоминает особенности обучающих данных и плохо обобщается на новые данные
- В) Модель использует слишком мало признаков для обучения

Правильный ответ: Б

Вопрос 3

Какой метод используется для автоматического подбора гиперпараметров в scikit-learn?

- А) GridSearchCV — перебор всех комбинаций гиперпараметров с кросс-валидацией
- Б) AutoML — встроенный метод автоматического подбора
- В) ManualSearch — ручной перебор без автоматизации

Правильный ответ: А

Вопрос 4

Что такое кросс-валидация (cross-validation)?

- А) Проверка модели только на тестовом наборе данных
- Б) Метод разбиения данных на несколько частей для более надёжной оценки качества модели
- В) Преобразование всех данных в одинаковый масштаб

Правильный ответ: Б

Вопрос 5

Какая стратегия помогает найти баланс между переобучением и недообучением?

- А) Использование только очень сложных моделей
- Б) Регуляризация модели, кросс-валидация и подбор гиперпараметров
- В) Использование максимально большого обучающего набора без проверки

Правильный ответ: Б

### Тема 13. «Ансамбли моделей»

Вопрос 1

Что такое ансамбль моделей в машинном обучении?

- А) Одна мощная модель, которая обучена на больших данных
- Б) Комбинация нескольких моделей, решения которых объединяются для улучшения качества предсказания
- В) Последовательное обучение одной модели несколько раз на одних данных

Правильный ответ: Б

Вопрос 2

Что такое баггинг (bootstrap aggregating)?

- А) Метод объединения нескольких моделей через взвешенное голосование
- Б) Обучение нескольких копий модели на разных подвыборках данных (с возвращением) и усреднение их предсказаний
- В) Метод, где каждая новая модель исправляет ошибки предыдущей

Правильный ответ: Б

Вопрос 3

Какой алгоритм использует идею случайного леса (Random Forest)?

- А) Один наиболее мощный алгоритм для всех задач
- Б) Ансамбль множества деревьев решений, каждое из которых обучено на случайной подвыборке данных и признаков
- В) Линейная комбинация нескольких линейных моделей

Правильный ответ: Б

Вопрос 4

Что такое бустинг (boosting) в машинном обучении?

- А) Случайное увеличение размера обучающего набора

Б) Метод, где каждая новая модель обучается с акцентом на ошибках, допущенных предыдущими моделями

В) Применение одной модели несколько раз без изменений

Правильный ответ: Б

Вопрос 5

Какое преимущество имеют ансамбли моделей перед одиночными моделями?

А) Требуют меньше времени на обучение и предсказание

Б) Часто имеют более высокое качество предсказания и меньше подвержены переобучению благодаря разнообразию моделей

В) Всегда гарантируют 100% точность

Правильный ответ: Б

#### **Тема 14. «Нейронные сети»**

Вопрос 1

Что такое искусственный нейрон (перцептрон)?

А) Аналог математической функции, состоящей из входов, весов, сумматора и функции активации

Б) Биологическая клетка в мозге человека

В) Алгоритм кластеризации данных

Правильный ответ: А

Вопрос 2

Какова роль функции активации в нейронной сети?

А) Функция активации нужна только для увеличения скорости обучения

Б) Вносит нелинейность в модель, позволяя сети обучаться на сложных примерах

В) Нормализует входные данные перед обработкой

Правильный ответ: Б

Вопрос 3

Что такое слои в нейронной сети?

А) Горизонтальные разделения данных на группы

Б) Уровни организации нейронов: входной слой, скрытые слои, выходной слой

В) Временные шаги при обработке последовательных данных

Правильный ответ: Б

Вопрос 4

Какой алгоритм используется для обучения нейронных сетей?

- А) Метод случайного поиска без направления
- Б) Градиентный спуск (backpropagation) — пересчёт весов на основе ошибки предсказания
- В) Метод полного перебора всех весов

Правильный ответ: Б

Вопрос 5

Какие примеры функций активации часто используются в нейронных сетях?

- А) Только линейная функция  $f(x) = x$
- Б) ReLU, сигмоида (sigmoid), tanh и другие нелинейные функции
- В) Только шаговая функция

Правильный ответ: Б

## Кейсы

### 1. Кейс по теме «Основные термины и понятия открытых данных.»

**Цель:** через практическую ситуацию закрепить термины «открытые данные», «набор данных», «машиночитаемый формат», «открытая лицензия», «персональные данные», «анонимизация» и др.

#### **Проблемная ситуация (АПК).**

Региональное министерство сельского хозяйства запускает портал открытых данных АПК. Планируется опубликовать:

- реестр сельскохозяйственных производителей (с указанием вида деятельности, площади земель, типов продукции);
- статистику урожайности по районам и культурам за последние 10 лет;
- данные о применении удобрений и пестицидов по хозяйствам в агрегированном виде;
- сведения о господдержке (субсидии по направлениям, без раскрытия ФИО фермеров).

Студенты должны предложить идеи ИИ-сервисов (прогноз урожайности, оценка рисков, рекомендации по посевным структурам) и определить, какие из перечисленных наборов могут быть опубликованы как открытые данные и в каком виде.

#### **Задания студентам.**

1. Идентификация наборов открытых данных АПК

Студенты выбирают, какие наборы могут стать открытыми при соблюдении условий:

- реестр производителей: опубликовать в виде открытого набора, если это юрлица/ИП и указаны только общедоступные реквизиты (ИНН/ОГРН, адрес хозяйства, виды продукции), без персональных данных физических лиц;
- статистика урожайности: может быть открытой, если данные агрегированы по району/культуре и не позволяют идентифицировать конкретное фермерское хозяйство;
- применение удобрений и пестицидов: публикация в агрегированном виде по району, типу культуры и классу вещества;
- господдержка: суммы по программам и категориям получателей в разрезе районов без указания ФИО фермеров.

Студенты указывают:

- какие из этих наборов явно являются «наборами открытых данных», если размещены в машиночитаемом формате (CSV, JSON) и снабжены открытой лицензией;
- какие наборы требуют анонимизации или агрегирования для исключения утечки персональных данных.

## 2. Персональные данные и анонимизация в АПК

Студенты определяют, где потенциально возникают персональные данные:

- фермеры-физлица (ФИО, адрес личного хозяйства, паспортные данные, персональные телефоны);
- данные о господдержке, если можно однозначно связать сумму с конкретным фермером.

Необходимо предложить способы анонимизации:

- замена ФИО на код/идентификатор или публикация только агрегированных данных по группам;
- укрупнение географии (район вместо конкретного населенного пункта), если малая численность получателей может привести к повторной идентификации.

## 3. Форматы и лицензии

Студенты формулируют рекомендации:

- в каких машиночитаемых форматах публиковать данные АПК (CSV для табличных данных, GeoJSON/SHAPE для пространственных данных, JSON/XML для API-доступа);
- какие условия должна содержать открытая лицензия (право на свободное использование, переработку и распространение при указании источника, отсутствие ограничений по сфере использования).

Отдельно обсуждается, почему без явной открытой лицензии даже технически доступные данные АПК нельзя считать полноценно «открытыми».

### **Результаты работы.**

По итогам работы студенты:

- заполняют таблицу с колонками: «Набор данных АПК», «Можно ли опубликовать как открытые данные?», «Требуемая анонимизация/агрегация», «Рекомендуемый формат», «Лицензионные условия»;

- формулируют своими словами термины «открытые данные», «набор открытых данных АПК», «машиночитаемый формат», «открытая лицензия», «персональные данные», «анонимизация» применительно к сельскому хозяйству.

## **2. Кейс по теме «Построение рядов распределения в Excel, Python. Описательные статистики в Excel, Python»**

Анализ выборки объектов недвижимости: студенты строят ряды распределения и вычисляют характеристики выборки в Excel и Python для ключевых количественных признаков (цены, площади, количества комнат).

### **Проблемная ситуация.**

Есть датасет квартир в крупном городе (например, выгрузка с портала объявлений):

- price — цена, тыс. руб.;
- area — общая площадь, м<sup>2</sup>;
- rooms — число комнат;
- floor, floors\_total, district и др.

**Цель:** подготовить аналитическую записку для агентства недвижимости:

- построить ряды распределения по количеству комнат и по цене;
- вычислить среднюю, моду и медиану цены и площади;
- сравнить результаты Excel и Python и интерпретировать различия.

### **Задачи.**

1. Ряды распределения в Python

Студенты:

- загружают датасет в pandas;
- строят негруппированный ряд распределения по rooms (частоты для 1, 2, 3, 4+ комнат);
- строят интервальный ряд по price (например, 0–3000, 3000–6000 тыс. руб. и т.д.);
- добавляют относительные частоты (деление частот на объем выборки).

По area студенты также формируют интервалы (до 30 м<sup>2</sup>, 30–50, 50–70, 70+ м<sup>2</sup>), чтобы увидеть структуру предложения по площади.

## 2. Ряд распределения и статистики в Excel

Студенты:

- выгружают те же данные (или подвыборку) в Excel;
- строят частотную таблицу по rooms (считать частоты по уникальным значениям);
- формируют интервальный ряд по price (либо вручную, либо через гистограмму/«Анализ данных»);
- считают:
  - среднюю цену (функция среднего);
  - медиану цены;
  - моду цены (или модальное значение по интервалам).

Аналогично — средняя, медиана и (при необходимости) мода по площади area.

## 3. Описательные статистики в Python

Студенты:

- вычисляют mean, median, mode для price и area через методы pandas или функции numpy/statistics;
- убеждаются, что результаты совпадают (с точностью до округления) с Excel, при условии одинаковой выборки и учета пропусков;
- строят простую визуализацию (гистограмму цены или площади) для наглядного сравнения с таблицей распределения.

Обсуждаются эффекты: вытянутость распределения цены вправо из-за дорогих объектов, положение медианы относительно средней, возможная многомодальность по количеству комнат (например, пики у 1- и 2-комнатных).

**Ожидаемый результат:** итоговый мини-отчет.

В отчете студенты представляют:

- таблицу ряда распределения по rooms и по интервалам price (частоты и доли);
- значения средней, медианы и моды цены и площади (с указанием, в какой среде посчитаны — Excel/Python);
- короткий текстовый вывод:
  - какая мера (средняя или медиана) лучше описывает «типичную» цену квартиры при наличии дорогих выбросов;
  - как структура по количеству комнат отражает профиль рынка (например, доминирование 1–2-комнатных квартир).

## 3. Кейс «Прогнозирование цен недвижимости» по теме «Машинное обучение»

**Цель:** применение линейной регрессии для построения прогностической модели.

**Проблемная ситуация.** Вы работаете аналитиком в агентстве недвижимости. Руководству необходимо предсказать цены квартир на основе площади и количества комнат.

**Задачи:**

1. Загрузить данные (площадь, количество комнат, цена).
2. Разделить их на обучающую и тестовую выборки.
3. Реализовать простую линейную регрессию с использованием библиотеки `scikit-learn`.
4. Построить графики зависимости фактических и предсказанных значений с помощью `matplotlib` или `plotly`.
5. Рассчитать метрики ошибки (MAE, MSE,  $R^2$ ).

**Ожидаемый результат:** обучающая программа, прогнозирующая стоимость квартиры по введенным характеристикам.

#### **4. Кейс «Классификация студентов по успеваемости» по теме «Машинное обучение»**

**Цель:** освоение базовых принципов классификации и метрик качества модели.

**Проблемная ситуация.** Отдел учебной аналитики университета хочет определить, какие студенты находятся в группе риска по результатам успеваемости.

**Задачи:**

1. Использовать набор данных о студентах (оценки, посещаемость, выполнение домашних заданий).
2. Реализовать бинарную классификацию с использованием алгоритма логистической регрессии или дерева решений.
3. Разделить данные на обучающую и тестовую выборки.
4. Построить `confusion matrix` и вычислить метрики точности (`accuracy`, `precision`, `recall`, `F1-score`).

**Ожидаемый результат:** модель, определяющая вероятность попадания студента в группу риска.

#### **5. Кейс «Сегментация клиентов интернет-магазина» по теме «Машинное обучение»**

**Цель:** изучение методов кластеризации без учителя.

**Проблемная ситуация.** Интернет-магазин хочет разделить клиентов на группы по поведению (средний чек, частота покупок, время последней активности).

**Задачи:**

1. Подготовить и нормализовать данные клиентов.
2. Реализовать метод `K-Means` и подобрать оптимальное число кластеров (метод локтя).
3. Визуализировать результаты с использованием двумерного графика кластеров.

4. Описать поведенческие особенности каждой найденной группы клиентов.

**Ожидаемый результат:** выявленные кластеры клиентов и их характеристика для маркетингового анализа.

## 6. Кейс «Классификация методом машинного обучения «Дерево решений»»

**Цель:** кейс ориентирован на практическое знакомство с классификацией методом дерева решений: постановка задачи, подготовка данных, обучение модели и интерпретация правил «если... то...» на реальном примере.

### Проблемная ситуация.

Компания по аренде недвижимости хочет автоматически классифицировать заявки клиентов на два класса:

- «высокий риск отказа» (клиент, скорее всего, откажется от сделки или не пройдет скоринг);
- «низкий риск отказа» (клиент с высокой вероятностью завершит сделку).

Есть датасет заявок с признаками: возраст, уровень дохода, наличие просрочек по кредитам, длительность аренды в прошлом, число несовершеннолетних детей, тип занятости и целевая метка `target` (0 — низкий риск, 1 — высокий риск).

### Задачи:

1. Постановка задачи и подготовка данных
  - Определить тип задачи: бинарная классификация методом дерева решений.
  - Разделить выборку на обучающую и тестовую (например, 70/30).
  - Обсудить, какие признаки могут влиять на риск отказа и нужно ли их кодировать (категориальные признаки, пропуски).
2. Обучение дерева решений
  - Обучить модель дерева решений на обучающей выборке (в Python с использованием библиотеки машинного обучения).
  - Настроить базовые гиперпараметры: максимальная глубина, минимальное число объектов в листе, критерий качества (Gini/entropy).
  - Оценить качество на тестовой выборке с помощью доли верных ответов и, при необходимости, матрицы ошибок.
3. Интерпретация дерева и правил  
Студентам нужно:
  - визуализировать дерево или вывести текстовое представление основных ветвей;
  - извлечь несколько человекочитаемых правил вида:
    - если доход < порога и есть просрочки, то «высокий риск»;
    - если доход  $\geq$  порога и нет просрочек, то «низкий риск»;

- объяснить, почему дерево выбрало именно такие признаки и пороги в верхних узлах (наиболее информативные признаки).

#### 4. Анализ ограничений и улучшений

Обсудить:

- риск переобучения при слишком глубоком дереве;
- влияние дисбаланса классов (если «высокий риск» встречается редко);
- идеи по улучшению: ограничение глубины, минимальный размер листа, использование ансамблей (случайный лес, градиентный бустинг) как развитие темы.

#### **Ожидаемый результат:**

Студенты готовят краткий отчет:

- описание исходной задачи и признаков;
- метрики качества модели на тесте (ассигасу, при желании precision/recall);
- 3–5 ключевых правил дерева в человекочитаемом виде и их интерпретация для бизнеса;
- вывод о пригодности дерева решений для автоматической предварительной оценки заявок и о рисках его использования.

### **6. Кейс по теме «Нейронные сети».**

**Цель:** базовое знакомство с нейронными сетями: от постановки задачи и подготовки данных до обучения простой модели и интерпретации результатов в практическом сценарии.

#### **Проблемная ситуация.**

Финтех-стартап хочет предсказывать вероятность отклика клиента на предложение по новой банковской карте. У компании есть датасет исторических маркетинговых кампаний с признаками: возраст, доход, количество продуктов в банке, наличие мобильного банка, количество обращений в колл-центр, участие в прошлых акциях и целевая метка `responded` (1 — откликнулся, 0 — нет).

Задача для студентов: построить простую нейронную сеть для бинарной классификации (`responded`) и сравнить ее работу с логистической регрессией или решающим деревом.

#### **Задачи:**

1. Постановка задачи и подготовка данных.

Студенты:

- формулируют задачу как задачу бинарной классификации по табличным данным;
- делят выборку на обучающую и тестовую (например, 70/30);
- выполняют предобработку:
  - нормализацию числовых признаков (масштабирование);

- кодирование категориальных признаков (one-hot);
- проверку и обработку пропусков.

Обсуждается, почему нейросети чувствительны к масштабу признаков и корректному кодированию категориальных переменных.

## 2. Архитектура простой нейросети.

Студенты проектируют базовую архитектуру:

- входной слой: размерность = число признаков после кодирования;
- 1–2 скрытых слоя (например, 16–32 нейрона) с нелинейной активацией (ReLU);
- выходной слой из 1 нейрона с сигмоидой для предсказания вероятности отклика.

Выбираются:

- функция потерь — бинарная кросс-энтропия;
- оптимизатор — простой градиентный метод (например, Adam);
- метрика качества — доля верных ответов, дополнительно по возможности F1 или AUC.

## 3. Обучение и оценка модели.

Студенты:

- обучают нейронную сеть на обучающей выборке с мини-батчами и несколькими эпохами;
- фиксируют динамику функции потерь и качества на обучающей и валидационной выборках (overfitting/underfitting);
- оценивают качество на тестовой выборке, сравнивая результаты с более простой моделью (логистическая регрессия или дерево решений).

Обсуждается:

- даёт ли нейросеть выигрыш по качеству по сравнению с простой моделью;
- как ограничение числа слоёв и нейронов влияет на переобучение и время обучения.

## 4. Интерпретация и практический вывод.

Студенты:

- анализируют, какие группы клиентов чаще всего получают предсказание «откликнется» (например, возраст 25–40, активный мобильный банк, 2–3 продукта);
- строят простые визуализации (распределение предсказанных вероятностей, ROC-кривую при наличии инструментария);
- формируют рекомендации для маркетинга:
  - какие сегменты целевой аудитории приоритетны;
  - на каких признаках модель, судя по анализу, «фокусируется» (по косвенным признакам, важности признаков из простой модели и т.п.).

**Ожидаемый результат:**

От группы ожидается мини-отчёт, включающий:

- постановку задачи и описание признакового пространства;
- архитектуру нейронной сети (слои, активации, размерности) и использованные гиперпараметры;
- сравнительную таблицу метрик для нейросети и базовой модели;
- интерпретацию результатов на языке предметной области (для руководства стартапа, а не для ML-специалистов).

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

Текущий контроль знаний, умений и навыков проводится в форме контрольной работы с расчетными задачами (или в форме теста), 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.

Максимальная оценка за выполнение каждой работы составляет 10 баллов.

Задержка выполнения индивидуального практического задания на одну неделю штрафуется одним баллом. Критерии оценки выполненных работ приведены в таблице:

<b>Оценка</b>	<b>Критерии оценивания</b>
«10» (отлично)	оценку «отлично» заслуживает студент, выполнивший практическую работу полностью, в работе корректно применены статистические методы, нет ошибок в расчетах, сделаны глубокие выводы. Студент дал полные ответы на все заданные вопросы по работе. Недостатков по оформлению работы не имеется.
«8-9» (хорошо)	оценку «хорошо» заслуживает студент, выполнивший практическую работу полностью, имеются недочеты в применении статистических методов, проведенном анализе и полученных выводах. Студент дал верные ответы на все заданные вопросы по работе. Недостатков по оформлению работы не имеется.
«6-7» (удовлетворительно)	оценку «удовлетворительно» заслуживает студент, выполнивший практическую работу полностью, но имеются существенные недочеты в применении отдельных статистических методов и полученных выводах. Студент дал верные ответы не на все вопросы. По оформлению работы имеются

	недостатки.
«1-5» (неудовлетворительно)	оценку «неудовлетворительно» заслуживает студент, выполнивший практическую работу не по своему варианту или с грубейшими нарушениями применения методов и последовательности анализа.

Контрольная работа оценивается по 100-бальной системе.

Оценка «отлично» выставляется студенту, если удельный вес правильных ответов составляет более 84%

Оценка «хорошо» выставляется студенту, если удельный вес правильных ответов составляет 75 – 84%

Оценка «удовлетворительно» выставляется студенту, если удельный вес правильных ответов составляет 60 – 74%

Оценка «неудовлетворительно» выставляется студенту, если удельный вес правильных ответов составляет более 0 – 59%

Итоговая оценка учитывает результаты рейтинговой системы контроля знаний (вклад 80%), результаты контрольных работ по всем темам (вклад 20%)

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. Трофимов, В. В. Алгоритмизация и программирование : учебник для вузов / В. В. Трофимов, Т. А. Павловская. — 4-е изд. — Москва : Издательство Юрайт, 2025. — 108 с. — (Высшее образование). — ISBN 978-5-534-20430-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/562040> (дата обращения: 15.08.2025).
2. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662> (дата обращения: 07.08.2025).
3. Демичев, В. В. Алгоритмизация и программирование: Учебное пособие / В. В. Демичев, Д. В. Быков, Д. Э. Храмов [и др.]; рец. С.Г. Сальников; Российский государственный аграрный университет - МСХА имени К.А. Тимирязева (Москва). — Электрон. текстовые дан. — Москва, 2024. — 248 с. — Коллекция: Учебная и учебно-методическая литература. — Свободный

- доступ из сети Интернет (чтение, печать, копирование). — Режим доступа : [http://elib.timacad.ru/dl/full/s17122024AP\\_v3.pdf](http://elib.timacad.ru/dl/full/s17122024AP_v3.pdf). - Загл. с титул. экрана. -Электрон. версия печ. публикации. — <URL:[http://elib.timacad.ru/dl/full/s17122024AP\\_v3.pdf](http://elib.timacad.ru/dl/full/s17122024AP_v3.pdf)>.
4. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 250 с. — (Высшее образование). — ISBN 978-5-534-20734-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558664> (дата обращения: 01.06.2025).
  5. Федоров, Д. Ю. Программирование на python : учебное пособие для вузов / Д. Ю. Федоров. — 6-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 187 с. — (Высшее образование). — ISBN 978-5-534-19666-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/556864> (дата обращения: 15.08.2025).
  6. Кремер, Н. Ш. Математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — 5-е изд. — Москва : Издательство Юрайт, 2025. — 259 с. — (Высшее образование). — ISBN 978-5-534-01654-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561039> (дата обращения: 05.08.2025).
  7. Статистика : учебник и практикум для вузов / под редакцией И. И. Елисеевой. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 380 с. — (Высшее образование). — ISBN 978-5-534-19581-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559668> (дата обращения: 14.07.2025).

## 7.2 Дополнительная литература

1. Зыков, С. В. Программирование : учебник и практикум для вузов / С. В. Зыков. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 285 с. — (Высшее образование). — ISBN 978-5-534-16031-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560815> (дата обращения: 10.08.2025).
2. Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики : пер. с англ. / под ред. Ю. В. Кирютенко. — 2-е изд. — М. : Вильямс, 2009. — 784 с. — ISBN 978-5-8459-1588-6.
3. Чернышев, С. А. Основы программирования на Python : учебник для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567821> (дата обращения: 17.08.2025).

4. Кормен, Т. Х. Алгоритмы: построение и анализ / Томас Х. Кормен, Чарльз Э. Лейзерсон, Рональд Л. Ривест, Клиффорд Штайн; пер. с англ. под ред. И. В. Красикова. — 2-е изд. — М. : Вильямс, 2012. — 1296 с. — ISBN 978-5-8459-1794-2.
5. Скиена, С. С. Алгоритмы. Руководство по разработке / Стивен С. Скиена ; пер. с англ. А. Л. Семёнова; под ред. А. К. Звонкова. — 2-е изд. — М.: ДМК Пресс, 2011. — 720 с. — ISBN 978-5-94074-714-0.
6. Стивенс, Род. Алгоритмы. Теория и практическое применение: [численные алгоритмы, структуры данных, методы работы с массивами, связанными списками и сетями] / Род Стивенс; [пер.: Кириленко Вадим, Волошко Роман Владимирович]. — Москва: Э, 2016. — 542, с. : ил., табл.; 24 см. — (Мировой компьютерный бестселлер). — ISBN 978-5-699-81729-0.
7. Бхаргава А. Грокаем алгоритмы. Иллюстрированное пособие для программистов и любопытствующих = Grokking Algorithms / пер. с англ. А. В. Белова. — Санкт-Петербург: Питер, 2017. — 288 с. : ил. — ISBN 978-5-496-02513-9.
8. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для вузов / Л. А. Станкевич. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 478 с. — (Высшее образование). — ISBN 978-5-534-20363-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560754> (дата обращения: 01.07.2025).
9. Кудрявцев, В. Б. Интеллектуальные системы : учебник и практикум для вузов / В. Б. Кудрявцев, Э. Э. Гасанов, А. С. Подколзин. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 165 с. — (Высшее образование). — ISBN 978-5-534-07779-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561954> (дата обращения: 02.08.2025).
10. Navarro G., Nekrich Y. Top-k Document Retrieval in Compressed Space // Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). — Philadelphia: SIAM, 2025. — P. 4009–4030. — DOI: 10.1137/1.9781611978322.137.
11. Fraser, K. C., Dawkins, H., & Kiritchenko, S. (2025). Detecting AI-generated text: Factors influencing detectability with current methods. *Journal of Artificial Intelligence Research*, 82, 2233–2278. <https://doi.org/10.1613/jair.1.16665>
12. Cheng S.-W., Huang H. Fréchet Distance in Subquadratic Time // Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). — Philadelphia : SIAM, 2025. — P. 5100–5113. — DOI: 10.1137/1.9781611978322.173.

13. Ellert J., Gawrychowski P., Górkiewicz A., Starikovskaya T. Faster two-dimensional pattern matching with k mismatches // Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). – Philadelphia : SIAM, 2025. – P. 4031–4060. – DOI: 10.1137/1.9781611978322.138.
14. Sowa, K., & Przegalińska, A. (2025). From expert systems to generative artificial experts: A new concept for human-AI collaboration in knowledge work. *Journal of Artificial Intelligence Research*, 82, 2103–2145. <https://doi.org/10.1613/jair.1.17175>
15. Valencia-Arias, A., Uribe-Bedoya, H., Peña, A., González-Ruiz, J. D., Sánchez Santos, G., Chapoñan Ramírez, E., & Martínez Rojas, E. (2024). Artificial intelligence and recommender systems in e-commerce: Trends and research agenda. *Intelligent Systems with Applications*, 24, 200435. <https://doi.org/10.1016/j.iswa.2024.200435>

### **7.3 Статьи, опубликованные в научных журналах 1 уровня Белого списка научных журналов Минобрнауки России и сборниках научных работ конференций уровня А\***

1. Bonifati, A., Voigt, H. Special issue on big graph data management and processing. *The VLDB Journal* 31, 201–202 (2022). <https://doi.org/10.1007/s00778-022-00732-6>

2. Liu, S., Yao, Y., Jia, J. et al. Rethinking machine unlearning for large language models. *Nat Mach Intell* 7, 181–194 (2025). <https://doi.org/10.1038/s42256-025-00985-0/>

3. Mittal, S., Thakral, K., Singh, R. et al. On responsible machine learning datasets emphasizing fairness, privacy and regulatory norms with examples in biometrics and healthcare. *Nat Mach Intell* 6, 936–949 (2024). <https://doi.org/10.1038/s42256-024-00874-y/>

4. Alistarh D., Kurtic E., Malinovsky G., Modoranu I.-V., Richtárik P., Robert T., Safaryan M. MicroAdam: Accurate Adaptive Optimization with Low Space Complexity // *Advances in Neural Information Processing Systems 37: Proc. of the 37th Conf. on Neural Information Processing Systems (NeurIPS 2024, Vancouver, Canada, 10–15 Dec. 2024)*. – Neural Information Processing Systems Foundation, Inc., 2024. – P. 1–43. – DOI: 10.52202/079017-0001.

5. Kang M., Park Y., Song C. SPO: Sequential Monte Carlo Policy Optimisation // *Advances in Neural Information Processing Systems 37: Proc. of the 37th Conf. on Neural Information Processing Systems (NeurIPS 2024, Vancouver, Canada, 10–15 Dec. 2024)*. – Neural Information Processing Systems Foundation, Inc., 2024. – P. 73–131. – DOI: 10.52202/079017-0003.

6. Гарбук С. В. Метод декомпозиции функциональных характеристик систем искусственного интеллекта // Искусственный интеллект и принятие решений. – 2025. – № 1. – С. 14–32. – DOI: 10.14357/20718594250102.

7. Yan J., Zhang W.-G., Liu Y., Pan W., Hou X.-Y., Liu Z.-Y. An autonomous navigation method for field phenotyping robot based on ground-air collaboration // Artificial Intelligence in Agriculture. – 2025. – Vol. 15, No. 4. – P. 610–621. – DOI: 10.1016/j.aiia.2025.05.005.

## **7.5 Методические указания, рекомендации и другие материалы к занятиям**

1. Кудрина, Е. В. Основы алгоритмизации и программирования на языке С# : учебное пособие для вузов / Е. В. Кудрина, М. В. Огнева. — Москва : Издательство Юрайт, 2022. — 322 с. — (Высшее образование). — ISBN 978-5-534-09796-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/517285> (дата обращения: 18.08.2022).
2. Кочегурова, Е. А. Теория и методы оптимизации : учебное пособие для вузов / Е. А. Кочегурова. — Москва : Издательство Юрайт, 2022. — 133 с. — (Высшее образование). — ISBN 978-5-534-10090-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490136> (дата обращения: 18.08.2022).

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

1. Официальный сайт Python. URL: <https://www.python.org/>.
2. Официальный сайт дистрибутива языков программирования Python и R Anaconda. URL: <https://www.anaconda.com/>.
3. Официальный сайт Росстата. URL: <https://rosstat.gov.ru/>.
4. Портал открытых данных Российской Федерации (<https://data.gov.ru>).
5. Портал открытых данных Правительства Москвы (<https://data.mos.ru>).
6. Каталог каталогов открытых данных (<https://www.datacatalogs.ru>).
7. Минфин России (<https://minfin.gov.ru/ru/opendata/>).
8. Единая точка доступа к открытым данным учреждений ЕС European data portal (<https://data.europa.eu>).
9. Open data portals Европейской комиссии (<https://digital-strategy.ec.europa.eu/en/policies/open-data-portals>).
10. Портал открытых данных правительства США Data.gov (<https://data.gov>).
11. Всемирный каталог открытых данных DataPortals.org (<https://dataportals.org>).
12. База данных о 2600+ порталах открытых данных в мире Open Data Inception (<https://opendatainception.io>).

13. Подборка международных открытых данных Open data resources (UK Data Service) (<https://ukdataservice.ac.uk/help/other-data-providers/open-data-resources/>).

## 9. Перечень программного обеспечения и информационных справочных систем

Таблица 9

### Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Практическая работа 1. «Основные термины и понятия открытых данных».	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Mi-crosoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
2	Практическая работа 2. «Автоматизированные способы получения открытых данных»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Mi-crosoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
3	Практическая работа 3. «Построение рядов распределения в Excel, Python».	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Mi-crosoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
4	Практическая работа 4. «Описательные статистики: средняя, мода, медиана в Excel, Python»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Mi-crosoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
5	Практическая работа 5. «Показатели	VS Code/Excel/Word/Anaconda (или свободно-	Расчетная/система управления па-	MMi-crosoft/Microsoft/Anaconda Inc. (или	2025/2007/2012 и позднее

	вариации в Excel, Python»	распространяемые аналоги)	кетами и дистрибутив	opensource)	
6	Практическая работа 6. «Загрузка данных в Python, работа с типом Series. работа с Data Frame»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
7	Практическая работа 7. «Библиотеки для визуализации данных в Python»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
8	Практическая работа 8. «Построение моделей корреляционно-регрессионного анализа»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
9	Практическая работа 9. «Классификация методом машинного обучения «Дерво решений»»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
10	Практическая работа 10. «Другие методы машинного обучения. Алгоритм KNN.»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
11	Практическая работа 11. «Другие методы машинного обучения. Кластер-	VS Code/Excel/Word/Anaconda (или свободно-распространяе-	Кроссплатформенный текстовый редактор для	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее

	ный анализ. Алгоритм KMeans.»	мые аналоги)	разработчиков/Расчетная/система управления пакетами и дистрибутив		
12	Практическая работа 12. «Оптимизация моделей, подбор параметров»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
13	Практическая работа 13. «Ансамбли моделей»	Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее
14	Практическая работа 14. «Нейронные сети»	VS Code/Excel/Word/Anaconda (или свободно-распространяемые аналоги)	Кроссплатформенный текстовый редактор для разработчиков/Расчетная/система управления пакетами и дистрибутив	Microsoft/Microsoft/Anaconda Inc. (или opensource)	2025/2007/2012 и позднее

### 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
учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и	Количество рабочих мест: 16 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450

<p><i>промежуточной аттестации (2й учебный корпус, 102 ауд.)</i></p>	<p>Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение: Институт Экономики и управления, Кафедра Статистики и кибернетики</p>
<p><i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (2й учебный корпус, 106 ауд.)</i></p>	<p>Корпус 2, Аудитория 106 Количество рабочих мест: 16 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение: Институт Экономики и управления, Кафедра Статистики и кибернетики</p>
<p><i>учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), 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></p>	<p>Корпус 2, Аудитория 302 Количество рабочих мест: 16 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение: Институт Экономик</p>
<p><i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (1й учебный корпус, 212 ауд.)</i></p>	<p>Корпус 1, Аудитория 212 Количество рабочих мест: 24 Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение: Кафедра Цифровая кафедра</p>

<p><i>учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий семинарского типа, учебная аудитория для проведения курсового проектирования (выполнения курсовых работ), учебная аудитория для групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации (1й учебный корпус, 214 ауд.)</i></p>	<p>Корпус 1, Аудитория 214          Количество рабочих мест: 24          Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro,          стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO,          питание PoE. Структурное подразделение: Кафедра          Цифровая кафедра</p>
<p><i>Центральная научная библиотека имени Н.И. Железнова</i></p>	<p>Читальные залы библиотеки</p>
<p><i>Студенческое общежитие</i></p>	<p>Комната для самоподготовки</p>

## **11. Методические рекомендации студентам по освоению дисциплины**

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

В ходе занятий вести конспектирование учебного материала. Обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. Задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

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

При подготовке к зачету (в конце семестра) повторять пройденный материал в строгом соответствии с учебной программой. Использовать конспекты и литературу, рекомендованную преподавателем. Обратит особое внимание на темы учебных занятий, пропущенных студентом по разным причинам. При необходимости обратиться за консультацией и методической помощью к преподавателю.

### **Виды и формы отработки пропущенных занятий**

Студент, пропустивший занятия обязан самостоятельно подготовиться к теме устного опроса, которые состоялись на практическом занятии. В рамках

часов консультаций студент может ответить на вопросы пропущенного устного опроса, которые были пропущены.

## **12. Методические рекомендации преподавателям по организации обучения по дисциплине**

Курс «Технологии работы с открытыми данными» должен давать не абстрактно-формальные, а прикладные знания. Данная цель может быть реализована только при условии соблюдения в учебных планах преемственности учебных дисциплин. Базовые знания для изучения Алгоритмизация и программирование дают такие дисциплины, как иностранный язык, экономическая теория, эконометрика. Освоение основных тем данной дисциплины позволит студентам сформировать представление о таком сложном предмете как макроэкономика, понять всю ширину науки и получить необходимые знания для последующего профессионального развития в этой области.

Студент может подготовить доклад по теме, представляющей его научный интерес, представить результаты в виде презентации. В случае надлежащего качества, его работа может быть заслушана на научном кружке кафедры или на студенческой научной конференции. По решению кафедры, студенты, занявшие призовые места на научных студенческих конференциях, могут освобождаться от сдачи зачета по этой дисциплине.

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

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

### **Программу разработал(и):**

Калитвин В.А., канд. ф.-м. наук, доцент

  
(подпись)

## РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.14 «Технологии работы с открытыми данными» ОПОП ВО по направлению 09.03.03 Прикладная информатика, направленность «Программные решения для бизнеса», (квалификация выпускника – бакалавр)

Иноземцева Алексея Ивановича, доцента кафедры высшей математики, кандидата физико-математических наук (далее по тексту рецензент), проведено рецензирование рабочей программы дисциплины «Технологии работы с открытыми данными» ОПОП ВО по направлению 09.03.03 Прикладная информатика, направленность «Программные решения для бизнеса» (бакалавриат) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре статистики и кибернетики (работчик - Калитвин Владимир Анатольевич, доцент, кандидат физико-математических наук).

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

1. Предъявленная рабочая программа дисциплины «Технологии работы с открытыми данными» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 09.03.03 Прикладная информатика. 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 – Б1.В.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 09.03.03 Прикладная информатика.

4. В соответствии с Программой за дисциплиной «Технологии работы с открытыми данными» закреплена 2 универсальные компетенции и 6 профессиональных компетенций. Дисциплина «Технологии работы с открытыми данными» 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 «Технологии работы с открытыми данными» составляет 3 зачётных единиц (108 часов).

7. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Технологии работы с открытыми данными» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.03.03 Прикладная информатика 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 «Технологии работы с открытыми данными» 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 09.03.03 Прикладная информатика.

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

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета в шестом семестре, что соответствует статусу дисциплины, как дисциплины части, формируемой участниками образовательных отношений – Б1.В ФГОС ВО направления 09.03.03 Прикладная информатика.

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

13. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 7 источников (базовые учебники), дополнительной литературой – 15 наименований, Интернет-ресурсы – 3 источника и соответствует требованиям ФГОС ВО направления *09.03.03 Прикладная информатика*.

14. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Технологии работы с открытыми данными» 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 «Технологии работы с открытыми данными».

#### **ОБЩИЕ ВЫВОДЫ**

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Технологии работы с открытыми данными» ОПОП ВО по направлению *09.03.03 Прикладная информатика*, направленности «Программные решения для бизнеса» (квалификация выпускника – бакалавр), разработанная Калитвиным Владимиром Анатольевичем, доцентом, кандидатом физико-математических наук, соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент:

Иноземцев Алексей Иванович, доцент кафедры высшей математики, кандидат физико-математических наук



«28» августа 2025 г.

(подпись)