

Автономная некоммерческая образовательная организация высшего образования «Институт цифровых профессий»

УТВЕРЖДАЮ
Ректор АНОО ВО
«Институт цифровых профессий»
Устименко В.А.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.12 Основы науки о данных

основной профессиональной образовательной программы
высшего образования – программы бакалавриата
по направлению подготовки 38.03.05 Бизнес-информатика,
направленность (профиль): Анализ данных в бизнесе

Квалификация: бакалавр

Форма обучения: очная

(с применением исключительно электронного обучения, дистанционных образовательных технологий)

Москва 2024 г.

ОГЛАВЛЕНИЕ

1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОЧЕЙ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ Б1.В.12 ОСНОВЫ НАУКИ О ДАННЫХ	3
2. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ	11
3. УСЛОВИЯ РЕАЛИЗАЦИИ УЧЕБНОЙ ДИСЦИПЛИНЫ	44
4. КОНТРОЛЬ И ОЦЕНКА РЕЗУЛЬТАТОВ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ	49

1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОЧЕЙ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.12 ОСНОВЫ НАУКИ О ДАННЫХ

1.1. Место дисциплины в структуре основной профессиональной образовательной программы

Учебная дисциплина Б1.В.12 Основы науки о данных входит в часть, формируемую участниками образовательных отношений, блока “Дисциплины” основной образовательной программы и реализуется в 1 и 2 семестрах.

Учебная дисциплина Б1.В.12 Основы науки о данных связана с дисциплинами: Анализ данных, Практикум по анализу данных, Введение в машинное обучение.

Образовательная программа реализуется с применением исключительно электронного обучения, дистанционных образовательных технологий.

1.2. Цель и планируемые результаты освоения дисциплины

Цель дисциплины – сформировать у студентов базовое понимание и практические навыки в области анализа данных, включая сбор, обработку, моделирование и визуализацию данных, для решения комплексных бизнес-задач и принятия обоснованных управленческих решений.

В результате обучения у студента формируются следующие компетенции:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный	УК-1.И-1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи. УК-1.И-2. Разрабатывает варианты решения проблемной ситуации	Знания: <ul style="list-style-type: none">● Понимает основные этапы Data Science и методологии CRISP-DM.● Знает типы данных и методов их обработки в Python.

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
<p>подход для решения поставленных задач.</p>	<p>на основе критического анализа доступных источников информации. УК-1.И-3. Выбирает оптимальный вариант решения задачи, аргументируя свой выбор. УК-1.И-4. Строит логические умозаключения на основе поступающих информации и данных для решения управленческих задач. УК-1.И-5. Выдвигает и обосновывает новые идеи, обосновывает альтернативы решения задач с учетом организационных, ресурсных, финансовых и иных ограничений и т.п.</p>	<ul style="list-style-type: none"> ● Понимает принципы моделирования и оценки моделей машинного обучения. <p>Умения:</p> <ul style="list-style-type: none"> ● Разрабатывает решения для анализа данных с использованием различных инструментов. ● Анализирует данные критически, выявляя аномалии, ошибки и ключевые закономерности. ● Синтезирует информацию для определения оптимальных решений на основе многокомпонентного анализа данных. <p>Навыки:</p> <ul style="list-style-type: none"> ● Критически оценивает доступные источники информации, включая данные из табличных процессоров, BI-инструментов и данные, полученные через API, для обоснования выбранных подходов и разработки различных сценариев решения проблемных ситуаций.

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
		<ul style="list-style-type: none"> ● Использует библиотеки и инструменты для поиска информации в открытых и специализированных источниках, применяя методологию CRISP-DM ● Генерирует новые идеи и подходы к решению задач на основе анализа данных, применяя инструменты предобработки, моделирования и визуализации данных, а также учитывать ограничения и ресурсы при разработке альтернативных решений.
<p>ОПК-4. Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для</p>	<p>ОПК-4.И-1. Осуществляет анализ и моделирование информационных потоков организации. ОПК-4.И-2. Использует методы и программные средства для сбора, обработки и анализа бизнес-информации. ОПК-4.И-3. Обладает навыками формирования и обоснования ИТ-решения для информационно-аналит</p>	<p>Знания:</p> <ul style="list-style-type: none"> ● Понимает основные концепции и принципы работы информационных технологий. ● Знает принципы работы с базами данных и SQL. ● Понимает методы сбора, обработки и анализа данных для поддержки управленческих решений. <p>Умения:</p>

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
информационно-аналитической поддержки принятия управленческих решений.	ической поддержки принятия управленческих решений.	<ul style="list-style-type: none"> ● Анализирует и моделировать информационные потоки организации с использованием специализированных инструментов. ● Использует методы и программные средства для эффективного сбора, обработки и анализа бизнес-информации. ● Формирует и обосновывать ИТ-решения для информационно-аналитической поддержки принятия управленческих решений. Навыки: <ul style="list-style-type: none"> ● Применяет инструменты для анализа и моделирования информационных потоков, используя методологии и инструменты, такие как ER-диаграммы для визуализации данных и оптимизации бизнес-процессов. ● Использует программные средства, такие как SQL для работы с базами данных, Python для обработки данных, и

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
		<p>BI-инструменты для визуализации данных, поддерживающих бизнес-аналитику.</p> <ul style="list-style-type: none"> ● Разрабатывает и обосновывает ИТ-решения, учитывая потребности организации и используя подходы, представленные в программе курса, для информационно-аналитической поддержки управленческих решений.
<p>ПК-5. Способен осуществлять анализ данных цифрового следа в соответствии с моделью деятельности человека (группы людей) и информационных систем.</p>	<p>ПК-5.И-1. Осуществляет подготовку данных, проводит оценку качества и очистку данных для проведения анализа, в том числе с использованием специализированного программного обеспечения</p> <p>ПК-5.И-2. Осуществляет проверку гипотез и поиск закономерностей</p> <p>ПК-5.И-3. Выполняет визуализацию данных и предлагает комплексные заключения по результатам обработки данных</p>	<p>Знания:</p> <ul style="list-style-type: none"> ● Понимает методы и инструменты для подготовки данных, включая оценку качества и очистку данных. ● Знает основы проверки гипотез и выявления закономерностей в данных. ● Понимает методы визуализации данных и способы представления аналитических выводов. <p>Умения:</p> <ul style="list-style-type: none"> ● Подготавливать данные для анализа, включая оценку их качества и очистку с использованием

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
		<p>специализированного программного обеспечения.</p> <ul style="list-style-type: none"> ● Проводить проверку гипотез и выявлять закономерности в данных с использованием статистических и аналитических методов. ● Создавать визуализации данных и формулировать комплексные выводы на основе анализа данных. <p>Навыки:</p> <ul style="list-style-type: none"> ● Использует инструменты и библиотеки для очистки и подготовки данных, а также проводит оценку их качества. ● Применяет методы проверки гипотез и анализа закономерностей в данных с использованием статистических методов и библиотек Python. ● Использует инструменты визуализации данных, для создания графиков и диаграмм, а также предлагать комплексные заключения

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
		по результатам анализа данных.
ПК-6. Способен управлять сбором и обработкой данных.	<p>ПК-6.И-1. Проектирует процесс сбора данных и определяет источники данных для проведения аналитических работ</p> <p>ПК-6.И-2. Осуществляет организацию работ по сбору и обработке данных</p> <p>ПК-6.И-3. Контролирует соответствие процесса получения и обработки данных заданному алгоритму.</p>	<p>Знания:</p> <ul style="list-style-type: none"> ● Знает принципы проектирования процессов сбора данных. ● Знает методы управления проектами и организации сбора данных. ● Знает алгоритмы и процедуры для контроля качества и соответствия данных. <p>Умения:</p> <ul style="list-style-type: none"> ● Проектирует процессы сбора данных, определяет источники данных и методы их получения. ● Организует и управляет работами по сбору и обработке данных, обеспечивая их эффективное выполнение. ● Контролирует процессы получения и обработки данных, проверяет их соответствие заданным алгоритмам и критериям качества. <p>Навыки:</p>

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Образовательные результаты
		<ul style="list-style-type: none"> ● Использует специализированные инструменты для проектирования процессов сбора данных и определения источников данных, таких как ER-диаграммы и схемы потоков данных. ● Применяет программные средства для организации и управления сбором данных, включая системы автоматизации сбора данных и управления проектами. ● Осуществлять мониторинг и контроль качества данных с помощью инструментов и алгоритмов проверки данных, таких как системы контроля версий и средства аудита данных.

2. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

2.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем в часах
В 3.Е.	8
Объем учебной дисциплины	288
в том числе:	
<i>Контактная работа</i>	102
- лекционного типа	68
- семинарского типа	34
<i>Самостоятельная работа</i>	150
Промежуточная аттестация: зачет (1 семестр)	-
Промежуточная аттестация: экзамен (2 семестр)	36

Образовательный процесс по дисциплине осуществляется с применением исключительно электронного обучения, дистанционных образовательных технологий с помощью электронной информационно-образовательной среды - образовательной платформы <https://go.skillbox.ru/>.

2.2. Тематический план и содержание учебной дисциплины Б1.В.12 Основы науки о данных

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
1 СЕМЕСТР					
Тема 1. Введение в науку о данных.	1.1 Понятие Data Science. Цели изучения Data Science. 1.2 Направления профессионального развития в области Data Science: дата аналитика, инженерия машинного обучения, дата инженерия. Роли в Data Science: общие задачи и различия. 1.3 Кейсы применения ML/AI в реальной жизни.	2	2	6	ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
Тема 2. Business Understanding. С чего начинается работа с данными.	2.1 Задачи Data Science с точки зрения бизнеса. 2.2 Понятие Data Mining. Методология CRISP-DM для исследования данных шесть фаз методологии. 2.3 Основные задачи Business Understanding. Выявление потребностей. Сбор требований и их виды.	2	-	6	ОПК-4, УК-1
Тема 3. Data Understanding. Использование табличного процессора.	3.1 Понятие Data Understanding. 3.2 Использование табличного процессора для обработки данных. Простые	2	-	6	ОПК-4, УК-1

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>меры для проверки качества данных.</p> <p>3.3 Статистические и логические функции.</p> <p>Сводные таблицы и графики.</p> <p>3.4 Как работают сводные таблицы. Объединение данных из нескольких таблиц.</p> <p>Выбор ПО для работы с таблицами.</p>				
Тема 4. Введение в Python.	<p>4.1 Python для анализа данных и Data Science, преимущества использования.</p> <p>Распространенность и версии Python.</p>	2	2	10	ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	4.2 Установка дистрибутива Anaconda и окружения. Облачные версии ПО. Запуск и работа в Jupyter Notebook. 4.3 Базовые вычисления на Python. Вывод на экран.				
Тема 5. Переменные и типы данных.	5.1 Стандартные типы данных: числа, строки. 5.2 Функции для чисел и строк. Варианты и причины ошибок. Справка с информацией по функциям. 5.3 Создание и применение переменной. Функции type и print. 5.4 Изменение переменной	4	2	8	УК-1, ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	на основе старого значения. Ввод данных с экрана. 5.5 Применение функции input. Создание нескольких input в одной ячейке. Несколько строк кода в ячейке. Булевы значения.				
Тема 6. Условия	6.1 Условные операторы if, else. Усложнение условия. Отступы, стандартный отступ. 6.2 Вложенные условия, оператор elif. 6.3 Дополнительные возможности, добавление условий в программу.	4	-	8	УК-1, ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	6.4 Ошибки и исключения. Разведение ошибок для понимания их источника.				
Тема 7. Циклы	7.1 Цикл for. Цикл while. Операторы continue, pass. Прерывание цикла, оператор break. 7.2 Бесконечный цикл. Логический тип данных 7.3 Цикл while со счетчиком. Оператор continue в цикле while. 7.4 Функция range. Функция range с началом отсчета. 7.5 Функция range: start, stop, step. Отрицательный шаг в	4	-	10	ОПК-4, УК-1

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	функции range. 7.6 Типовые алгоритмы со счетными циклами.				
Тема 8. Алгоритмы и структуры данных	8.1 Алгоритмы, эффективность и сложность. Операции над структурами данных. 8.2 Списки в Python. Функции и методы для списков. Цикл while, условие else и while true для списков. Операции со списками: сортировка, slicing. 8.3 Линейный и бинарный поиск, преимущества и бинарного поиска.	4	2	10	ОПК-4, УК-1

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	8.4 Выделение индекса элемента по значению. 8.5 Виды сортировок. Сортировка пузырьком.				
Тема 9. Функции	9.1 Определение и классификация функций. Объявление и тестирование функций. Требования к наименованию функции. Сохранение результата функций в разные переменные. Возвращаемые значения. 8.2 Понятие и реализация декомпозиции. Удобство работы со списками в	4	2	10	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	Python. 8.3 Разбор практической задачи: расчет NPS. Методы и функции. Встроенные функции. Три способа вывода значений на экран. Обращение к индексам. 8.4 Нормализация строк. Понятие рекурсия.				
Тема 10. Коллекции в Python	10.1 Коллекции: кортежи, словари и множества. Создание пустого кортежа и заполнение значениями. Добавление и переопределение элемента. 10.2 Понятие Unicode.	4	-	6	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>Понятие множества. Работа со встроенными функциями. Методы для работы с множествами. Поиск пересечений, объединение двух множеств. Операции над множествами.</p> <p>10.3 Понятие словарей. Два способа создания словаря. Добавление ключей, удаление лишней пары.</p> <p>10.4 Генерация списков на основе других списков.</p> <p>10.5 Руководство по оформлению кода PEP 8.</p>				

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
Тема 11. Чтение файлов в Python и командной строке.	<p>11.1 Чтение файлов в Python. Основные операции с файлами. Понятие объекта. Последствия незакрытых файлов. Применение менеджера контекста в программе.</p> <p>11.2 Официальная документация.</p> <p>11.3 Роль и автономия командной строки. Запуск терминала, выполнение базовых команд. Создание одной директории внутри другой. Работа с файлами в командной строке.</p>	2	2	6	УК-1, ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>11.4 Где живет Python. Пример проекта в Python. Повторное использование кода.</p> <p>11.5 Работа с командой pip и ее подкомандами. Установка пакетов с помощью pip.</p>				
Тема 12. Библиотека Pandas.	<p>12.1 Понятие pandas. Чтение файлов с помощью pandas. Основы чтения данных из различных форматов (CSV, Excel).</p> <p>12.2 JSON-файлы. Чтение и обработка JSON-файлов.</p> <p>12.3 Подсчет простых</p>	2	-	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>статистик в Pandas. Вычисление базовых статистических показателей.</p> <p>12.4 Объединение датафреймов. Методы объединения данных из разных источников.</p> <p>12.5 Применение изученных методов на практике.</p>				
Промежуточная аттестация (зачет)				-	
Всего за 1 семестр		36	18	90	
2 СЕМЕСТР					
Тема 13. Получение данных с помощью API	13.1 Инструменты API, REST API. Основные понятия и принципы работы API и REST API.	2	-	4	УК-1, ОПК-4

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>13.2 HTTP-запросы с помощью Postman. Использование Postman для отправки HTTP-запросов.</p> <p>13.3 Requests. Работа с библиотекой Requests для выполнения HTTP-запросов в Python.</p> <p>13.4 Подготовка файлов для передачи другим разработчикам. Документирование файлов.</p> <p>13.5 Установка и настройка Git. Основы работы с системами контроля версий.</p>				

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
Тема 14. Базы данных	<p>14.1 Понятие баз данных. Реляционная модель данных. Примеры лучших СУБД.</p> <p>14.2 Основные концепции и типы баз данных. Методы подключения к различным базам данных.</p> <p>14.3 Создание и использование ER-диаграмм для моделирования данных. ER-модель, типы связей в ER-модели.</p> <p>14.4 Нормализация баз данных. Принципы нормализации данных.</p> <p>14.5 Различия между OLTP и</p>	2	-	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	OLAP системами. 14.6 Применение изученных методов на практике.				
Тема 15. Язык запросов SQL	15.1 Введение в SQL. Основные понятия SQL и типы данных. Базовые операции SQL. Копирование запроса в буфер обмена. 15.2 Основные команды для работы с данными. Использование функций агрегирования и группировки данных. 15.3 Методы объединения данных из нескольких таблиц.	2	2	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	Операторы INNER JOIN, LEFT JOIN, RIGHT JOIN, UNION. 15.4 Применение изученных методов на практике: написание SQL-запросов для доступа к данным.				
Темы 16. Визуализация данных с помощью BI-инструментов .	16.1 BI-инструменты. Обзор и преимущества использования BI-инструментов. 16.2 Процесс установки инструмента. Загрузка данных в BI-инструмент. 16.3 Основы визуализации в инструментах бизнес-аналитики.	2	2	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>16.4 Добавление и визуализации данных. Добавление элементов интерактивности.</p> <p>16.5 Подключение и использование данных из различных источников. Список инструментов для визуализации данных.</p> <p>16.6 Применение изученных методов на практике.</p>				
Тема 17. Подготовка данных	17.1 Введение в тему подготовки данных. Data Preparation как одна из фаз CRISP-DM. Основные цели и задачи подготовки данных.	2	-	4	УК-1, ОПК-4, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>источники данных, неструктурированный формат данных.</p> <p>17.2 Определение и работа с целевой переменной. Поиск целевых переменных под конкретную задачу.</p> <p>17.3 Различие и работа с дискретными и непрерывными переменными. Группы переменных: количественные и качественные.</p> <p>17.4 Распределения вероятностей. Анализ и визуализация</p>				

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	распределений данных. 17.5 Применение изученных методов на практике.				
Тема 18. Разведочный анализ данных. Data Cleaning	18.1 Понятие разведочного анализа данных. 18.2 Проверка на дубликаты, пропуски, типы данных, выбросы и зависимости. Как проверить дублирующие строки. Поиск пропущенных значений. Неинформативные данные. Приведение типов. Корректировка типов данных. Поиск аномалий и нахождение выбросов	2	2	4	УК-1, ОПК-4, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	18.3 Применение изученных методов на практике.				
Тема 19. Разведочный анализ данных: Data Visualization	19.1 Введение в визуализацию данных. Примеры удачной и неудачной визуализации данных. Путь построения графика от полотна до подписей. 19.2 Основы визуализации данных с использованием библиотеки matplotlib. Построение гистограммы. Методы визуализации для обнаружения аномалий. Поиск аномалий с помощью	2	2	4	УК-1, ОПК-4, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>boxplot.</p> <p>19.3 Визуализация категориальных переменных. Агрегирующие функции для групп. Визуализация и анализ корреляций между переменными. Корреляция между всеми численными переменными.</p> <p>19.4 Применение изученных методов на практике: работать с аномалиями в данных, строить визуализации, а также выявлять зависимости между</p>				

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	переменными с помощью библиотеки Matplotlib.				
Тема 20. Разведочный анализ данных. Feature Engineering	20.1 Понятие feature engineering. Проверка связи признаков. Использование lambda функций для создания новых признаков. 20.2 Генерация дополнительных строковых фичей. Методы генерации новых строковых признаков. Преобразование категориальных данных для использования в моделях. 20.3 Добавление новых признаков и данных из	2	2	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	матрицы. 20.4 Нормализация данных. Методы и цели нормализации данных. Проверка зависимости до и после стандартизации. Обзор ключевых аспектов. 20.5 Применение изученных методов на практике.				
Тема 21. Моделирование .	21.1 Методология CRISP-DM: фаза Modeling. Основные понятия моделирования. 21.2 Создание моделей и постановка гипотез. Формулирование и проверка гипотез.	2	2	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	21.3 Основные статистические критерии. Парный Т-критерий и критерий Манна — Уитни. Знаковый ранговый тест Вилкоксона. 21.4 Применение изученных методов на практике.				
Тема 22. Машинное обучение.	22.1 Этап моделирования у инженера и аналитика. Роль моделирования в работе ML-инженера. 22.2 Основы машинного обучения. Виды задач машинного обучения: классификация, регрессия,	2	-	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>кластеризация.</p> <p>22.3 Основные метрики качества классификации. Выявление ошибок модели.</p> <p>22.4 Основы работы с деревьями решений. Формирование дерева решений.</p> <p>22.5 Шаги по обучению первой модели машинного обучения. Создание тренировочной и тестовой выборки.</p> <p>22.6 Случайный лес. Улучшение качества модели на тестовой выборке. Подбор</p>				

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	гиперпараметров моделей машинного обучения. 22.7 Применение изученных методов на практике.				
Тема 23. Линейные модели и нейронные сети.	23.1 Обзор темы линейных моделей и нейронных сетей. Линейные алгоритмы. Обучение модели линейной регрессии на примере задачи. 23.2 Логистическая регрессия. Обучение модели бинарной классификации. Проверка точности модели.	2	-	4	УК-1, ОПК-4, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>23.3 Нейронные сети. Слои, синапсы и связи. Фреймворки для разработки нейросетей.</p> <p>23.4 Применение изученных методов на практике.</p>				
Тема 24. Метрики в аналитике.	<p>24.1 Основные понятия метрик в аналитике. Продуктовые и маркетинговые метрики. Примеры целевых метрик.</p> <p>24.2 Основные правила построения дашборда.</p> <p>24.3 Определение и анализ метрик тщеславия. Перевод метрики тщеславия в</p>	2	-	4	УК-1, ПК-5

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формированию которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	здоровую метрику. Здоровые метрики: метрики активности пользователей, метрики удержания, конверсионные метрики.				
Тема 25. Маркетинговая и продуктовая аналитика.	25.1 Основы и задачи маркетинговой аналитики. Анализ, примеры и этапы воронок продаж. 25.2 Методы анализа эффективности интернет-магазинов. Маржа, средний чек и конверсия. Качество клиента, повышение ROI. 25.3 Основы и инструменты	2	-	4	УК-1, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	<p>продуктовой аналитики. Метрики роста и метрики продукта. Метрики эффективности решения задачи пользователя. 25.4 Проведение когортного анализа и сегментации. 25.5 Применение изученных методов на практике.</p>				
Тема 26. Modeling, заключение.	<p>26.1 Обзор заключительных тем моделирования. 26.2 Методы анализа категорий и доли рынка. Анализ мнений, сегментов и возможностей роста. 26.3 Применение изученных</p>	2	-	4	УК-1, ОПК-4, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	методов на практике: построение интерактивного дашборда.				
Тема 27. Оценка модели.	27.1 Оценка модели с точки зрения бизнеса. Способы оценки модели. 27.2 Переобучение и оценка результатов моделирования. Понятие переученной и недоученной кривой. 27.3 Понятие кросс-валидации. Подготовка кода модели к передаче. Проверка правильности сохранения модели. Создание	4	-	4	УК-1, ОПК-4, ПК-5, ПК-6

Наименование разделов и тем	Содержание учебного материала	Контактная работа с преподавателем		Самостоятельная работа (ак.ч.)	Коды компетенций, формирование которых способствует элемент программы
		Лекционного типа (ак.ч.)	Семинарского типа (ак.ч.)		
	предсказания с помощью модели. Список сериализуемых объектов. 27.4 Практическая работа: задача предсказания стоимости подержанного автомобиля.				
Промежуточная аттестация (экзамен)				36	
Всего за 2 семестр		32	16	96	
Итого по дисциплине: 288		68	34	186	

3. УСЛОВИЯ РЕАЛИЗАЦИИ УЧЕБНОЙ ДИСЦИПЛИНЫ

3.1. Для реализации образовательной программы с применением исключительно электронного обучения, дистанционных образовательных технологий в организации созданы условия функционирования электронной информационно-образовательной среды.

Электронная информационно-образовательная среда организации функционирует на образовательной платформе <https://go.skillbox.ru/>.

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронной информационно-образовательной среде организации из любой точки, в которой имеется доступ к информационно-телекоммуникационной сети «Интернет».

При реализации программы с применением исключительно электронного обучения, дистанционных образовательных технологий предполагается режим обучения, при котором обучающийся осваивает образовательную программу удаленно, взаимодействуя с педагогическим работником исключительно посредством цифровых образовательных сервисов и ресурсов электронной информационно-образовательной среды, и отсутствуют учебные занятия, проводимые путем непосредственного взаимодействия педагогического работника с обучающимся в аудитории.

Используются активные и интерактивные формы обучения и воспитания как в синхронном, так и в асинхронном режиме.

Синхронное электронное обучение с применением исключительно дистанционных образовательных технологий реализуется в форме проведения и участия в различного типа вебинарах и в одновременное работе в чате с преподавателем в рамках письменных онлайн-консультаций.

Асинхронное электронное обучение с применением исключительно электронного обучения, дистанционных образовательных технологий реализуется в части контактной работы в форме отложенной во времени обратной связи от преподавателя на сданные обучающимся на проверку работы, а также с самостоятельным изучением учебных текстовых и видеоматериалов в ЭИОС.

Общими принципами освоения обучающимися учебного материала с применением исключительно электронного обучения, дистанционных образовательных технологий являются:

- участие в вебинарах;
- самостоятельное изучение материалов лекций и других электронных учебных материалов;
- просмотр видео-лекций и записей вебинаров в ЭИОС;
- выполнение заданий и загрузка их на проверку преподавателю через личный кабинет ЭИОС;
- ознакомление с результатами оценивания качества выполнения работ и повторная загрузка на проверку при необходимости доработки через личный кабинет ЭИОС;
- выполнение самостоятельной работы в соответствии с тематическим планом дисциплины;
- участие в консультационных и вводных ознакомительных вебинарах с преподавателями;
- общение с преподавателями посредством чата в личном кабинете ЭИОС;
- получение в личном кабинете ЭИОС индивидуальных заданий от преподавателей;
- прохождение процедур текущего контроля успеваемости и промежуточной аттестации в личном кабинете ЭИОС;
- дистанционная техническая поддержка обучающихся по работе в личном кабинете ЭИОС;
- формирование обратной связи по качеству учебного материала, условиям обучения;
- формирование индивидуального электронного портфолио.

В учебном процессе используются активные и интерактивные методы, такие как вебинары, онлайн-консультации, проведение мини-исследований, написание эссе, разбор профессиональных ситуаций, решение ситуационных задач, анализ видеофрагментов, выполнение творческих заданий и другие, с предоставлением учебных материалов в электронном виде.

Используемое программное обеспечение (в т.ч. отечественного производства):

- а) Лицензионное ПО (в т.ч. отечественного производства):
 - не используется для освоения дисциплины.

б) Свободно распространяемое ПО (в т.ч. отечественного производства):

- онлайн-редактор документов в Облаке Mail.ru;
- Jupyter Notebook;
- Redash;
- Docker;
- GitLab.

3.2. Информационное обеспечение реализации программы

Для реализации программы библиотечный фонд образовательной организации имеет электронные учебные издания (электронные образовательные ресурсы) для использования в образовательном процессе.

3.2.1. Основные электронные учебные издания (электронные образовательные ресурсы)

1. Шелудько, В. М. Основы программирования на языке высокого уровня Python: учебное пособие / В. М. Шелудько; Южный федеральный университет. - Ростов-на-Дону; Таганрог: Издательство Южного федерального университета, 2017. - 146 с. - ISBN 978-5-9275-2649-9. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1021662>.

2. Мартишин, С. А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем: учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва: ФОРУМ: ИНФРА-М, 2024. — 368 с. — (Высшее образование). - ISBN 978-5-8199-0946-1. - Текст: электронный. - URL: <https://znanium.com/catalog/product/2096940>.

3. Агальцов, В. П. Базы данных: в 2 книгах. Книга 2. Распределенные и удаленные базы данных: учебник / В.П. Агальцов. — Москва: ФОРУМ: ИНФРА-М, 2024. — 271 с. — (Высшее образование). - ISBN 978-5-8199-0959-1. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2138458>.

3.2.2. Дополнительные источники

1. Лыгина, Н. И. Моделирование: учебное пособие / Н. И. Лыгина, О. В. Лауферман. - Новосибирск: Изд-во НГТУ, 2020. - 87 с. - ISBN 978-5-7782-4151-0. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1866931>.

2. Кулаичев, А. П. Методы и средства комплексного статистического анализа данных: учебное пособие / А.П. Кулаичев. — 5-е изд., перераб. и доп. — Москва: ИНФРА-М, 2022. — 484 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/25093. - ISBN 978-5-16-012834-4. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/1815604>.

3.3. Электронные информационные ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Перечень электронных информационных ресурсов, рекомендованных для освоения образовательной программы:

Наименование ресурса	Ссылка
Министерство науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
Федеральная служба по надзору в сфере образования и науки	https://obrnadzor.gov.ru/
Федеральный портал «Российское образование»	https://edu.ru/
Информационная система «Единое окно доступа к образовательным ресурсам»	http://window.edu.ru/
Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru/
Персональная образовательная платформа	eor-madk.com.ru

Наименование ресурса	Ссылка
Электронно-библиотечная система «Знаниум» (ресурс приспособлен для использования инвалидами и лицами с ограниченными возможностями здоровья)	znanium.com
Образовательная платформа «Юрайт»	https://urait.ru/
Информационный банковский портал	banki.ru
Электронный ресурс Банка России	cbr.ru
Электронно-библиотечная система Book.ru	https://www.book.ru
Образовательная платформа LearningApps	learningapps.org
Научная электронная библиотека eLIBRARY.RU	eLIBRARY.RU
Якласс	https://www.yaklass.ru
Учи.ру	https://uchi.ru
Образовательная онлайн-платформа Инфоурок	https://infourok.ru
Библиотека видеоуроков	https://interneturok.ru/
Портал о бизнес-планировании (на английском языке)	https://business-plany.ru/
Федеральная служба государственной статистики	https://rosstat.gov.ru/
Единая межведомственная информационно-статистическая система	https://www.fedstat.ru/
Документация по языку Python	https://docs.python.org/3/contents.html
Электронный ресурс для тех, кто делает сайты	https://htmlbook.ru/

4. КОНТРОЛЬ И ОЦЕНКА РЕЗУЛЬТАТОВ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

4.1. Результаты обучения

Знания:

- Знает основные этапы Data Science и методологию CRISP-DM.
- Знает типы данных и методы их обработки в Python.
- Знает принципы моделирования и оценки моделей машинного обучения.
 - Понимает основные концепции и принципы работы информационных технологий.
 - Знает принципы работы с базами данных и SQL.
 - Знает методы сбора, обработки и анализа данных для поддержки управленческих решений.
 - Знает методы и инструменты для подготовки данных, включая оценку качества и очистку данных.
 - Знает основы проверки гипотез и выявления закономерностей в данных.
 - Знает методы визуализации данных и способы представления аналитических выводов.
 - Знает принципы проектирования процессов сбора данных.
 - Знает методы управления проектами и организации сбора данных.
 - Знает алгоритмы и процедуры для контроля качества и соответствия данных.

Умения:

- Разрабатывает решения для анализа данных с использованием различных инструментов.
 - Анализирует данные критически, выявляя аномалии, ошибки и ключевые закономерности.
 - Синтезирует информацию для определения оптимальных решений на основе многокомпонентного анализа данных.
 - Анализирует и моделирует информационные потоки организации с использованием специализированных инструментов.
 - Использует методы и программные средства для эффективного сбора, обработки и анализа бизнес-информации.

- Формирует и обосновывает ИТ-решения для информационно-аналитической поддержки принятия управленческих решений.

- Подготавливает данные для анализа, включая оценку их качества и очистку с использованием специализированного программного обеспечения.

- Проводит проверку гипотез и выявляет закономерности в данных с использованием статистических и аналитических методов.

- Создает визуализации данных и формулирует комплексные выводы на основе анализа данных.

- Проектирует процессы сбора данных, определяет источники данных и методы их получения.

- Организовывает и управляет работами по сбору и обработке данных, обеспечивая их эффективное выполнение.

- Контролирует процессы получения и обработки данных, проверяет их соответствие заданным алгоритмам и критериям качества.

Навыки:

- Критически оценивает доступные источники информации, включая данные из табличных процессоров, BI-инструментов и данные, полученные через API, для обоснования выбранных подходов и разработки различных сценариев решения проблемных ситуаций.

- Использует библиотеки и инструменты для поиска информации в открытых и специализированных источниках, применяя методологию CRISP-DM.

- Генерирует новые идеи и подходы к решению задач на основе анализа данных, применяя инструменты предобработки, моделирования и визуализации данных, а также учитывает ограничения и ресурсы при разработке альтернативных решений.

- Применяет инструменты для анализа и моделирования информационных потоков, используя методологии и инструменты, такие как ER-диаграммы для визуализации данных и оптимизации бизнес-процессов.

- Использует программные средства, такие как SQL для работы с базами данных, Python для обработки данных, и BI-инструменты для визуализации данных, поддерживающих бизнес-аналитику.

- Разрабатывает и обосновывает ИТ-решения, учитывая потребности организации и используя подходы, представленные в

программе курса, для информационно-аналитической поддержки управленческих решений.

- Использует инструменты и библиотеки для очистки и подготовки данных, а также проводит оценку их качества.
- Применяет методы проверки гипотез и анализа закономерностей в данных с использованием статистических методов и библиотек Python.
- Использует инструменты визуализации данных для создания графиков и диаграмм, а также предлагает комплексные заключения по результатам анализа данных.
- Использует специализированные инструменты для проектирования процессов сбора данных и определения источников данных, таких как ER-диаграммы и схемы потоков данных.
- Применяет программные средства для организации и управления сбором данных, включая системы автоматизации сбора данных и управления проектами.
- Осуществлять мониторинг и контроль качества данных с помощью инструментов и алгоритмов проверки данных, таких как системы контроля версий и средства аудита данных.

4.2.1. Критерии оценки зачета

«Зачтено» - теоретическое содержание курса освоено полностью, на достаточном уровне сформированы умения, знания и навыки, все предусмотренные рабочей программой задания выполнены без существенных ошибок.

«Не зачтено» - теоретическое содержание курса не освоено или освоено не полностью, необходимые умения, знания и навыки не сформированы, выполненные задания содержат существенные ошибки.

4.2.2. Критерии оценки экзамена

«Отлично» - теоретическое содержание курса освоено полностью, высокий уровень сформированности умений, знаний и навыков, все предусмотренные рабочей программой задания выполнены без ошибок.

«Хорошо» - теоретическое содержание курса освоено полностью, хороший/средний уровень сформированности умений, знаний и

навыков, все предусмотренные рабочей программой задания выполнены, некоторые виды заданий выполнены с несущественными ошибками.

«Удовлетворительно» - теоретическое содержание курса в основном освоено, пороговый уровень сформированности умений, знаний и навыков, большинство предусмотренных рабочей программой заданий выполнено, некоторые из выполненных заданий содержат существенные ошибки.

«Неудовлетворительно» - теоретическое содержание курса не освоено или освоено не полностью, уровень сформированности умений, знаний и навыков ниже порогового, выполненные задания содержат существенные и критические ошибки.

4.3 Формы и методы контроля

Текущий контроль по дисциплине представляет собой проверку результатов освоения учебного материала по учебной дисциплине. Текущий контроль проводится преподавателем в ЭИОС в период всего освоения курса на основе оценочных средств по учебной дисциплине при использовании следующих форм исключительно электронного обучения, дистанционных образовательных технологий:

- синхронное и асинхронное наблюдение за ходом выполнения практических и лабораторных работ, решения обучающимися ситуационных, профессиональных и иных задач посредством ЭИОС;
- проверка в ЭИОС качества выполнения практических работ, в т.ч. подготовка и участие в обсуждении вопросов, вынесенных на семинары;
- оценка в ЭИОС результатов групповой работы обучающихся, в т.ч. в рамках деловых игр и групповых проектов;
- проверка в ЭИОС выполнения самостоятельной работы обучающихся (изучение, конспектирование, реферирование), в т.ч. письменных работ (эссе, статьи, презентации, схемы, таблицы, доклады, проекты, индивидуальные задания и других);
- оценка ответов обучающихся в рамках фронтальных или индивидуальных устных и/или письменных опросов в ЭИОС, в т.ч. в рамках семинарских занятий, вебинаров-тренингов, учебных бесед и других форм работы;

- проведение тестирования в ЭИОС по отдельным темам или разделам учебной дисциплины.

Промежуточная аттестация по дисциплине проходит в форме зачета (1 семестр) и в форме экзамена (2 семестр).

Промежуточная аттестация по дисциплине проводится преподавателем в ЭИОС после окончания освоения курса в период экзаменационной сессии на основе оценочных средств по учебной дисциплине при использовании исключительно электронного обучения, дистанционных образовательных технологий.