

**Федеральное государственное образовательное бюджетное  
учреждение высшего образования  
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ  
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»  
(Финансовый университет)**

**Кафедра искусственного интеллекта  
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

Проректор по учебной и  
методической работе

\_\_\_\_\_ Е.А. Каменева  
24.12.2024 г.

**Макрушин С.В., Блохин Н.В.**

**Технологии обработки данных**

**Рабочая программа дисциплины**

для студентов, обучающихся по направлению подготовки:  
09.03.03 - Прикладная информатика,  
ОП «Прикладные информационные системы в экономике и финансах»

*Рекомендовано Ученым советом  
Факультета информационных технологий и анализа больших данных  
(протокол № 50 от 17.12.2024 г.)*

*Одобрено заседанием Кафедры искусственного интеллекта  
(протокол № 5 от 04.12.2024 г.)*

**Москва 2024**

## Содержание

1. Наименование дисциплины.....	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.....	2
3. Место дисциплины в структуре образовательных программ.....	3
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся .....	3
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	4
5.1. Содержание дисциплины .....	4
5.2. Учебно-тематический план.....	7
5.3. Содержание семинаров, практических занятий .....	10
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	12
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы .....	12
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю .....	13
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	14
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины .....	17
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины .....	18
10. Методические указания для обучающихся по освоению дисциплины .	20
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем .....	21
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине. ....	21

## 1. Наименование дисциплины

«Технологии обработки данных».

## 2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПKN-6	Способность организовывать поиск и сбор информации, ее хранение в структурированном виде, проектировать и реализовывать реляционные и нереляционные базы и хранилища данных	1. Демонстрирует знание основ реляционных баз данных, нормализации данных, ACID, CRUD, ORM, использует транзакции.	<b>Знать:</b> способы хранения данных с отношениями «один ко многим» и «один к одному».  <b>Уметь:</b> с помощью языка Python сохранять и читать данные с отношениями «один ко многим» и «один к одному».
		2. Демонстрирует знание различных технологий хранения данных: реляционные и нереляционные базы данных, документарные хранилища, извлекает данные из разных источников и в разных форматах, в том числе программно.	<b>Знать:</b> различные технологии сериализации, хранения и чтения данных.  <b>Уметь:</b> с помощью языка Python сохранять и читать данные в популярных универсальных форматах.
		3. Проектирует хранилища данных исходя из их назначения и характера данных, выбирает инструментальное и архитектурное решение, физическую и логическую схему данных и обосновывает свой выбор.	<b>Знать:</b> подходы к организации структуры данных в форматах CSV, XML и JSON.  <b>Уметь:</b> с помощью языка Python сохранять и читать данные в форматах CSV, XML и JSON.

### 3. Место дисциплины в структуре образовательных программ

Дисциплина «Технологии обработки данных» относится к Общепрофессиональному циклу дисциплин по направлению подготовки 09.03.03 - Прикладная информатика, ОП «Прикладные информационные системы в экономике и финансах».

### 4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

#### *очная форма обучения*

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 2 (в часах)
<b>Общая трудоёмкость дисциплины</b>	<b>4/144</b>	<b>144</b>
<b>Контактная работа- Аудиторные занятия</b>	<b>50</b>	<b>50</b>
Лекции	16	16
Семинары, практические занятия	34	34
<b>Самостоятельная работа</b>	<b>94</b>	<b>94</b>
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

#### *очно-заочная форма обучения*

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 4 (в часах)
<b>Общая трудоёмкость дисциплины</b>	<b>4/144</b>	<b>144</b>
<b>Контактная работа- Аудиторные занятия</b>	<b>34</b>	<b>34</b>
Лекции	16	16
Семинары, практические занятия	18	18
<b>Самостоятельная работа</b>	<b>110</b>	<b>110</b>
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

*заочная форма обучения (ИОО)*

<b>Вид учебной работы по дисциплине</b>	<b>Всего (в з.е. и часах)</b>	<b>Семестр 2 (в часах)</b>
<b>Общая трудоёмкость дисциплины</b>	<b>4/144</b>	<b>144</b>
<b><i>Контактная работа- Аудиторные занятия</i></b>	<b>12</b>	<b>50</b>
Лекции	4	4
Семинары, практические занятия	8	8
<b><i>Самостоятельная работа</i></b>	<b>132</b>	<b>132</b>
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

**5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий**

**5.1. Содержание дисциплины**

**Тема 1. Библиотека NumPy и Pandas.**

В рамках темы рассматривается технологический стек Python для обработки и анализа данных, возможности Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры Маскирование и прихотливое индексирование в NumPy.

В рамках темы рассматриваются возможности библиотеки Pandas. Организация Pandas DataFrame и организация индексации для DataFrame и Series; применение универсальных функций и работа с пустыми значениями в Pandas.

Объединение данных из нескольких Pandas DataFrame: общая логика и примеры. Рассматривается операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение».

## **Тема 2. Использование различных форматов файлов в задачах обработки данных.**

В рамках темы рассматриваются принципы работы с файлами, файлы и операционные системы. Специфика текстовых и бинарных файлов.

В рамках темы рассматривается задача сериализации и десериализации данных и использование различных форматов файлов для ее решения. Описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python.

В рамках темы рассматриваются формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup.

В рамках темы рассматривается проблематика форматов файлов для хранения и обработки больших данных. Форматы файлов NPY и HDF: общая характеристика, пример взаимодействия с данными этих форматов в Python.

## **Тема 3. Взаимодействие с табличными данными в приложениях обработки данных.**

В рамках темы рассматривается формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python.

В рамках темы рассматриваются возможности использования Excel для внешних приложений обработки данных. Взаимодействие с Excel из Python с помощью библиотеки XLWings: принципы работы и примеры использования.

#### **Тема 4. Визуализация данных.**

В рамках темы рассматриваются основы работы с библиотекой `matplotlib`: организация системы координат, оформление осей, цвета и цветовые карты в `matplotlib`, стили линий и маркеры. `Pyplot` и объектно-ориентированный интерфейс `matplotlib`. Управление фигурами и создание множества графиков на одном рисунке. Различные типы графиков.

В рамках темы рассматривается визуализация данных с помощью библиотеки `Pandas`: набор методов для построения графиков, реализованный в структурах `Series` и `DataFrame`.

В рамках темы проводится введение в разведочный анализ данных: типы признаков, анализ распределений, анализ мер центральной тенденции и поиск выбросов, анализ взаимного распределения и парных корреляций. Проведение разведочного анализа данных с помощью библиотеки `Seaborn`.

#### **Тема 5. Работа со строками в приложениях обработки данных.**

В рамках темы рассматриваются возможности `python` по форматированию строк: `%`-форматирование, метод `format`, `f`-строки.

В рамках темы рассматриваются основы работы с регулярными выражениями: базовый синтаксис, примеры. Модуль `re` в `Python`. Примеры использования регулярных выражений.

В рамках темы рассматривается использования хэширования при работе со строками. Строки в библиотеке `pymru`.

#### **Тема 6. Введение в обработку текста на естественном языке в задачах обработки данных.**

В рамках темы рассматриваются сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на `Python`. Использование мемоизации на примере работы со строками. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на `Python`. Векторное представление текста на естественном языке: общий алгоритм подходов `TF`; `TF-IDF`.

## Тема 7. Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.

В рамках темы рассматривается профилирование реализации алгоритмов на Python, принципы решения задачи оптимизации производительности алгоритма. Библиотека Numba: принципы работы, базовые примеры использования. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции.

### 5.2. Учебно-тематический план

*очная форма обучения*

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа- Аудиторная работа			Самосто- ятельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практиче- ские занятия		
1	Библиотека NumPy и Pandas	24	10	4	6	14	Самостоя- тельные ра- боты. Участие в решении за- дач на прак- тических за- нятиях. Собе- седования по домашним за- даниям.  Самостоя- тельные ра- боты. Участие в решении за- дач на прак- тических за- нятиях. Собе- седования по домашним за- даниям.
2	Использование различных форма- тов файлов в зада- чах обработки данных.	20	6	2	4	14	
3	Взаимодействие с табличными дан- ными в приложе- ниях обработки данных.	22	8	2	6	14	
4	Визуализация данных	18	6	2	4	12	
5	Работа со стро- ками в приложе- ниях обработки данных	20	6	2	4	14	
6	Введение в обра- ботку текста на естественном	22	8	2	6	14	



	языке в задачах обработки данных						Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
7	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.	18	6	2	4	12	
	В целом по дисциплине	144	50	16	34	94	Согласно учебному плану: контрольная работа
	Итого в %		35	32	68	65	

*очно-заочная форма обучения*

№ п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа- Аудиторная работа			Самосто- ятельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практиче- ские занятия		
1	Библиотека NumPy и Pandas	24	8	4	4	16	Самостоя- тельные ра- боты. Участие в решении за- дач на прак- тических за- нятиях. Собе- седования по домашним за- даниям.
2	Использование различных форма- тов файлов в зада- чах обработки данных.	22	6	2	4	16	
3	Взаимодействие с табличными дан- ными в приложе- ниях обработки данных.	20	4	2	2	16	
4	Визуализация данных	20	4	2	2	16	
5	Работа со стро- ками в приложе- ниях обработки данных	20	4	2	2	16	
6	Введение в обра- ботку текста на	20	4	2	2	16	

	естественном языке в задачах обработки данных						Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
7	Профилирование процессов обработки данных, библиотека Numba и векторизация в NumPy и Numba.	18	4	2	2	14	
	В целом по дисциплине	144	34	16	18	110	Согласно учебному плану: контрольная работа
	Итого в %		24	47	53	76	

заочная форма обучения (ИОО)

№ п/п	Наименование темы (раздела) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	*Контактная работа- Аудиторная работа			Самосто- ятельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практиче- ские занятия		
1	Библиотека NumPy и Pandas	24	4	2	2	20	Самостоя- тельные ра- боты. Участие в решении за- дач на прак- тических за- нятиях. Собе- седования по домашним за- даниям.  Самостоя- тельные ра- боты. Участие в решении за- дач на прак-
2	Использование различных форматов файлов в задачах обработки данных.	24	4	2	2	20	
3	Взаимодействие с табличными данными в приложениях обработки данных.	22	2	0	2	20	
4	Визуализация данных	20	2	0	2	18	
5	Работа со строками в приложениях обработки данных	18	0	0	0	18	

6	Введение в обработку текста на естественном языке в задачах обработки данных	18	0	0	0	18	тических занятиях. Собеседования по домашним заданиям.
7	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba.	18	0	0	0	18	
	В целом по дисциплине	144	12	4	8	132	Согласно учебному плану: контрольная работа
	Итого в %		8	33	67	92	

\* объем контактной работы в очно-заочной/заочной формах обучения и индивидуальных учебных планах определяется соответствующими учебными планами. Темы, реализуемые в виде контактной работы, определяются преподавателем самостоятельно, исходя из уровня их сложности

### 5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Библиотека NumPy и Pandas	<ul style="list-style-type: none"> <li>• Технологический стек Python для обработки и анализа данных</li> <li>• Возможности Python как glue language</li> <li>• Организация массивов в NumPy: хранение данных, создание массивов</li> <li>• Принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy.</li> <li>• Организация Pandas DataFrame и организация индексации для DataFrame и Series.</li> <li>• Применение универсальных функций и работа с пустыми значениями в Pandas.</li> <li>• Объединение данных из нескольких Pandas DataFrame: общая логика и примеры.</li> </ul> 8[1], 9[9], 9[10]	Интерактивная форма, работа на компьютере

Использование различных форматов файлов в задачах обработки данных	<ul style="list-style-type: none"> <li>• Формат файлов Pickle, представление данных в этом формате и взаимодействие с ним в Python.</li> <li>• Формат файлов JSON, представление данных в этом формате и взаимодействие с ним в Python.</li> <li>• Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM</li> <li>• Работа с XML с помощью библиотеки BeautifulSoup.</li> </ul> 8[1], 8[2], 9[3], 9[4]	Интерактивная форма, работа на компьютере
Взаимодействие с табличными данными в приложениях обработки данных.	<ul style="list-style-type: none"> <li>• Взаимодействие с Excel из Python с помощью библиотеки XLWings.</li> <li>• Формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python</li> </ul> 8[1], 8[2]	Интерактивная форма, работа на компьютере
Визуализация данных	<ul style="list-style-type: none"> <li>• Построение визуализаций с помощью библиотеки matplotlib</li> <li>• Построение визуализаций с помощью библиотеки pandas</li> <li>• Построение визуализаций с помощью библиотеки seaborn</li> </ul> 8[1], 9[13], 9[15], 9[16]	Интерактивная форма, работа на компьютере
Работа со строками в приложениях обработки данных	<ul style="list-style-type: none"> <li>• Основы работы с регулярными выражениями: базовый синтаксис, примеры.</li> <li>• Модуль re в Python.</li> </ul> 8[1], 8[2], 9[4]	Интерактивная форма, работа на компьютере
Введение в обработку текста на естественном языке в задачах обработки данных.	<ul style="list-style-type: none"> <li>• Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python.</li> <li>• Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python.</li> </ul> 8[1], 8[2], 9[4], 9[5], 9[6]	Интерактивная форма, работа на компьютере
Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba	<ul style="list-style-type: none"> <li>• профилирование реализации алгоритмов на Python</li> <li>• принципы решения задачи оптимизации производительности алгоритма</li> <li>• Библиотека Numba: принципы работы, базовые примеры использования.</li> </ul> 8[1], 8[2], 9[1], 9[2], 9[3]	Интерактивная форма, работа на компьютере

## 6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

### 6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Библиотека NumPy и Pandas	<ul style="list-style-type: none"> <li>• Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры.</li> <li>• Маскирование и прихотливое индексирование в NumPy.</li> <li>• Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение».</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Использование различных форматов файлов в задачах обработки данных	<ul style="list-style-type: none"> <li>• Формат файлов NPY, представление данных в этом формате и взаимодействие с ним в Python.</li> <li>• Формат файлов HDF, представление данных в этом формате и взаимодействие с ним в Python.</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Взаимодействие с табличными данными в приложениях обработки данных.	<ul style="list-style-type: none"> <li>• Продвинутое взаимодействие с Excel из Python с помощью библиотеки XLWings.</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Визуализация данных	<ul style="list-style-type: none"> <li>• Построение трехмерных графиков</li> <li>• Продвинутое взаимодействие с цветовыми картами</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Работа со строками в приложениях обработки данных	<ul style="list-style-type: none"> <li>• Использование хэширования при работе со строками.</li> <li>• Строки в библиотеке numpy.</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.

Введение в обработку текста на естественном языке в задачах обработки данных.	<ul style="list-style-type: none"> <li>• Использование мемоизации на примере работы со строками.</li> <li>• Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF.</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Профилирование процессов обработки данных, библиотека Numba и векторизация в NumPy и Numba	<ul style="list-style-type: none"> <li>• Векторизация в numru: ключевые параметры функции, примеры применения</li> <li>• Использование обобщенной сигнатуры функции в numru и numba.</li> </ul>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.

## 6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

### *Примерные вопросы к контрольной работе*

1. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
2. Универсальные функции и применение функций по осям в NumPy
3. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
4. Маскирование и прихотливое индексирование в NumPy
5. Векторизация в numru: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции

### *Примерные задания контрольной работы*

1. В массиве чисел, хранящихся в файле `finance.csv` найти строку (вывести ее индекс и содержащиеся значения), в которой более всего значений, превышающих среднее значение по всему массиву. Для расчётов использовать Pandas.
2. В массиве чисел, хранящихся в файле `finance.csv`, подсчитать количество строк, в которых более 600 значений больше среднего значения по всему массиву. Для расчётов использовать Pandas.

3. В массиве чисел, хранящихся в файле `finance.csv`, подсчитать количество значений, не отклоняющихся от среднего значения более чем на 3 стандартных отклонения. Для расчетов использовать `Pandas`.

*Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры искусственного интеллекта Факультета информационных технологий и анализа больших данных.*

## **7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. **«Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».**

**Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний**

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения ( умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
<b>ПKN-6</b> Способность организовывать поиск и сбор информации, ее хранение в структурированном виде, проектировать и реализовывать реля-	1.Демонстрирует знание основ реляционных баз данных, нормализации данных, ACID, CRUD, ORM, использует транзакции.	<b>Знать:</b> способы хранения данных с отношениями «один ко многим» и «один к одному».  <b>Уметь:</b> с помощью языка Python сохранять и читать данные с отношениями «один ко многим» и «один к одному».	Сохранить информацию телефонной книги, содержащей отношения «один к одному» и «один ко многим» в формате XML.

ционные и нереляционные базы и хранилища данных	2. Демонстрирует знание различных технологий хранения данных: реляционные и нереляционные базы данных, документарные хранилища, извлекает данные из разных источников и в разных форматах, в том числе программно.	<p><b>Знать:</b> различные технологии сериализации, хранения и чтения данных.</p> <p><b>Уметь:</b> с помощью языка Python сохранять и читать данные в популярных универсальных форматах.</p>	Сериализовать информацию телефонной книги в формате Pickle и XML.
	3. Проектирует хранилища данных исходя из их назначения и характера данных, выбирает инструментальное и архитектурное решение, физическую и логическую схему данных и обосновывает свой выбор.	<p><b>Знать:</b> подходы к организации структуры данных в форматах CSV, XML и JSON.</p> <p><b>Уметь:</b> с помощью языка Python сохранять и читать данные в форматах CSV, XML и JSON.</p>	Сохранить и прочитать информацию телефонной книги используя формат JSON.

### ***Примерные вопросы для подготовки к экзамену***

1. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
2. Универсальные функции и применение функций по осям в NumPy
3. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
4. Маскирование и прихотливое индексирование в NumPy
5. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции
6. Numba: принципы работы, базовые примеры использования
7. Организация Pandas DataFrame и организация индексации для DataFrame и Series
8. Применение универсальных функций и работа с пустыми значениями в Pandas



9. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры
10. Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение»
11. Специфика текстовых и бинарных файлов, форматы файлов CSV и Pickle, представление данных в этих форматах и взаимодействие с ними в Python
12. Задача сериализации и десериализации, описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python
13. Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup
14. Форматы файлов NPY и HDF общая характеристика, пример взаимодействия с данными этих форматов в Python
15. Взаимодействие из Python с базой данных на примере API SQLite, базовые возможности работы с транзакциями
16. Взаимодействие с Excel из Python с помощью XLWings: принципы работы и примеры использования
17. Основы работы с регулярными выражениями: базовый синтаксис, примеры использования модуля re в Python
18. Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python
19. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python
20. Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF

### **Пример экзаменационного билета**

1. **(20 баллов)** Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными

2. **(20 баллов)** Датасет: Chinook\_Sqlite.sqlite

С помощью кода на Python с использованием sqlite3 и SQL решить задачу. Реализовать функции на Python:

1. Которая возвращает все имеющиеся плейлисты.
2. Которая по имени плейлиста возвращает количество треков в нем и их суммарную продолжительность.

3. **(20 баллов)** Датасет: sp500hst.txt

Создать таблицу, в которой индексом являются даты торгов, столбцами - наименования тикеров, а в ячейках хранятся объемы торгов. Заполнить эту таблицу данными из sp500hst.txt (в случае отсутствия информации для определенных сочетаний тикер-дата, сохранить в ячейке пустое значение). Сохранить результат в новый CSV файл. Решить задачу средствами numpy и/или pandas.

### **8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**

#### ***Основная литература:***

1. Колдаев, В. Д. Структуры и алгоритмы обработки данных: учебное пособие / В. Д. Колдаев. - Москва: РИОР: ИНФРА-М, 2021. - 296 с. - ЭБС ZNANIUM. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 18.11.2024). – Текст: электронный.

2. Коротеев, М. В. Основы машинного обучения на Python: учебник / М. В. Коротеев. — Москва: КноРус, 2024. — 431 с. — Текст: непосредственный. — То же. - ЭБС BOOK.ru. — URL: <https://book.ru/book/952751> (дата обращения: 18.11.2024). — Текст: электронный.

***Дополнительная литература:***

3. Нагаева, И. А. Основы алгоритмизации и программирования: практикум: учебное пособие / И. А. Нагаева, И. А. Кузнецов. — Москва, Берлин: Директ-Медиа, 2021. — 169 с. — ЭБС Университетская библиотека ONLINE. — URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 18.11.2024). — Текст: электронный.

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

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>

2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>

4. Электронно-библиотечная система Znanium <http://www.znanium.com>

5. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>

6. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>

7. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>

8. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>

9. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>

10. Математические журналы: полнотекстовая коллекция Математического института им. В.А. Стеклова РАН <https://www.mathnet.ru/>
11. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
12. Национальная электронная библиотека <http://нэб.рф/>
13. Ресурсы информационно-аналитического агентства по финансовым рынкам Cbonds.ru <https://cbonds.ru/>
14. СПАРК <https://spark-interfax.ru/>
15. CNKI. Academic Reference <https://ar.oversea.cnki.net/>
16. Электронные продукты издательства Elsevier <http://www.sciencedirect.com>
17. Emerald: Management eJournal Portfolio <https://www.emerald.com/insight/>
18. Реферативная база данных по математике MathSciNET <https://mathscinet.ams.org/mathscinet/>
19. Коллекция научных журналов Oxford University Press <https://academic.oup.com/journals/>
20. Электронные коллекции книг и журналов издательства Springer: <http://link.springer.com/>
21. Платформа STATISTA <https://www.statista.com/>
22. База данных научных журналов издательства Wiley <https://onlinelibrary.wiley.com/>
23. Pylru 1.0.9 [Электронный ресурс]: сайт. – Режим доступа: <https://pypi.python.org/pypi/pylru>
24. Python Data Analysis Library [Электронный ресурс]: сайт. – Режим доступа: <http://pandas.pydata.org/>
25. Python Documentation [Электронный ресурс]: сайт. – Режим доступа: <http://python.org/doc/>
26. Python Standard Library [Электронный ресурс]: сайт. – Режим доступа: <https://docs.python.org/2/library/>

27. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. – Режим доступа: <http://scikit-learn.org>
28. Официальный сайт продукта <https://www.python.org/>
29. Каталог курсов Интернет Университета Информационных Технологий <http://www.intuit.ru/>
30. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>
31. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>
32. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>

## **10. Методические указания для обучающихся по освоению дисциплины**

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет-источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, помечая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

## **11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем**

### **11.1. Комплект лицензионного программного обеспечения:**

1. Пакет офисных программ;
2. Антивирус Kaspersky;

### **11.2. Современные профессиональные базы данных и информационные справочные системы:**

1. Информационно-правовая система «Гарант»;
2. Информационно-правовая система «Консультант Плюс»;
3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>;
4. Система комплексного раскрытия информации «СКРИН» - <http://www.skrin.ru/>;

### **11.3. Сертифицированные программные и аппаратные средства защиты информации: - не используются.**

## **12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Для проведения лекций и практических занятий необходима аудитория, оснащенная проектором и компьютерами с постоянным подключением к сети Интернет.