

**Федеральное государственное образовательное
бюджетное учреждение высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент анализа данных и машинного обучения
Факультета информационных технологий и анализа больших данных**

СОГЛАСОВАНО

Газпромбанк

(наименование организации)

Начальник управления

(должность представителя работодателя)

_____ Дашковский В.А.

(подпись)

(ФИО)

20.12.2023 г.

УТВЕРЖДАЮ

Проректор по учебной

и методической работе

_____ Е.А. Каменева

28.12.2023 г.

С.В. Макрушин, Е.Ю. Ключков

Технологии работы с открытыми данными

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

01.03.02 - Прикладная математика и информатика,

ОП "Прикладное машинное обучение",

Профиль: «Прикладное машинное обучение»

Рекомендовано Ученым советом

*Факультета информационных технологий и анализа больших данных
(протокол №39 от 20.12.2023 г.)*

Одобрено Советом учебно-научного

*Департамента анализа данных и машинного обучения
(протокол №10 от 14.12.2023 г.)*

Москва 2023

Содержание

2

2

Ошибка! Закладка не определена.

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

Ошибка! Закладка не определена.

4

5.2. Учебно-тематический план.....6

7

6. Учебно-методическое обеспечение для самостоятельной работы обучающихся по дисциплине.....8

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....8

9

10

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....15

16

17

18

18

1. Наименование дисциплины

«Технологии работы с открытыми данными»

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
ПКП-3	Способность выполнять структурирование данных, выявлять признаки ценные для моделей машинного обучения	1. Демонстрирует знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.	Знать: Принципы процесса структурирования данных и формирования признаков для моделей машинного обучения. Уметь: Демонстрировать знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.
		2. Применяет инструментальные средства для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.	Знать: Как применять инструментальные средств для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения. Уметь: Применять инструментальные средств для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.
		3. Владеет практическим навыком инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.	Знать: Принципы инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи. Уметь: Использовать практические навыки инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.

ПКП-1	Способность собирать наборы данных, в том числе больших данных, выполнять их подготовку для анализа в соответствии с решаемой прикладной задачей	1. Владеет навыками поиска внешних и внутренних источников данных для решения прикладной задачи.	<u>Знать:</u> Где искать внешние и внутренние источники данных для решения прикладной задачи. <u>Уметь:</u> Использовать навыки поиска внешних и внутренних источников данных для решения прикладной задачи.
		2. Использует инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников.	<u>Знать:</u> Инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников. <u>Уметь:</u> Использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников.
		3. Владеет навыками обработки больших данных для решения прикладных задач машинного обучения.	<u>Знать:</u> Как использовать навыки обработки больших данных для решения прикладных задач машинного обучения. <u>Уметь:</u> Использовать навыки обработки больших данных для решения прикладных задач машинного обучения.

3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии работы с открытыми данными» является дисциплиной Профиля "Прикладное машинное обучение" по направлению подготовки 01.03.02 - Прикладная математика и информатика, ОП "Прикладное машинное обучение".

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

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7 (в часах)
Общая трудоемкость дисциплины	4 з.е. 144 ч.	144 ч.
Контактная работа – Аудиторные занятия	68	68
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	52	52
Самостоятельная работа	76	76
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

Тема1. Основные форматы данных. REST API. Запросы на динамический сайт.

- Основные форматы данных (JSON, XML, CSV, YAML, INI) и их применение.
- `GET`, `POST`, `PUT`, `PATCH` и `DELETE` - запросы на динамический сайт.

Тема 2. Командные запросы: "Джак запросы". Протоколы (HTTP, FTP, SMTP и т. д.) для Curl запросов. Способы объединения (Merge) данных из разных источников.

- Curl запросы для тестирования и проверки API,
- Curl запросы для получения доступа к удаленному ресурсу и
- Curl запросы для мониторинга работоспособности сайтов и веб-приложений.
- Настройка Curl запросов под конкретные потребности.

- Использование SQL-запросов, использование оператора "JOIN".
- Использование инструментов Business Intelligence (BI) - QlikView, Tableau, Power BI.
- Слияние данных из разных источников в единую таблицу. BI-системы
- Объединение данные из разных баз данных, файлов или API.
- Инструменты для интеграции данных - Talend, Pentaho, Informatica.

Тема 3. Парсинг веб-страниц (веб-скрапинг). Мультипроцессинг.

- Процесс автоматического извлечения информации со страниц веб-сайтов - **web scraping**.
- автоматизация процесса парсинга веб-страниц, использование библиотек 'BeautifulSoup', 'Scrapy', 'multiprocessing', 'concurrent.futures'.
- Использование нескольких процессов для выполнения различных задач одновременно на многоядерном или многопроцессорном компьютере.
- Модуль 'multiprocessing'.
- Модуль 'concurrent.futures'.

Тема 4. Получение данных с wiki / wikidata. Обход блокировок при парсинге сайта.

- API для Википедии. использование библиотеки 'wikipediaapi'.
- API для Wikidata, использование библиотеки 'pywikibot'.
- Методы обхода некоторых типов блокировок.
- Настраиваемые заголовки (Custom headers), библиотека 'requests', чтобы запрос выглядел, как будто его отправил браузер.
- Прокси-сервер (Proxy server) - перенаправление запросов с компьютера на сайт, для сокрытия ваш IP-адреса.
- Задержка (Delay), для равномерного распределения загрузки на сайт, для избежания блокировки.

5.2. Учебно-тематический план

№ п/ п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа*			Самостоя тельная работа	
			Общая, в т. ч.:	Лекции	Семинары, практическ ие занятия		
О ш иб ка ! Не ук аз ан а по сл ед ов ат ел ьн ос ть.	Основные форматы данных. REST API. Запросы на динамический сайт.	36	17	4	13	19	Самостоятельная работа. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
О ш иб ка ! Не ук аз ан а по сл ед ов ат ел ьн ос ть.	Командные запросы: "Джак запросы". Протоколы (HTTP, FTP, SMTP и т. д.) для Curl запросов. Способы объединения (Merge) данных из разных источников.	36	17	4	13	19	Самостоятельная работа. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.

О ш иб ка ! Не ук аз ан а по сл ед ов ат ел ьн ос ть.	Парсинг веб-страниц (веб-скрапинг). Мультипроцессинг.	36	17	4	13	19	Самостоятельная работа. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
О ш иб ка ! Не ук аз ан а по сл ед ов ат ел ьн ос ть.	Получение данных с wiki / wikidata. Обход блокировок при парсинге сайта.	36	17	4	13	19	Самостоятельная работа. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
	В целом по дисциплине	144	68	16	52	76	Согласно учебному плану: контрольная работа
	Итого в %		47	24	76	53	

* 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
Тема 1 Основные форматы данных. REST API. Запросы на динамический сайт.	Форматы данных - CSV, XML, JSON. Эмуляция работы браузера Обращение REST API к своему бэкэнду Использование библиотеки Requests. Использование библиотеки Selenium. <i>Рекомендуемая литература 8.1-8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 2. Командные запросы: "Джак запросы". Протоколы (HTTP, FTP, SMTP и т. д.) для Curl запросов. Способы объединения (Merge) данных из разных источников.	"Джак запросы" - Curl (Client URL) запросы, Командные запросы через командную строку, с использованием утилиты для передачи данных через различные протоколы (HTTP, FTP, SMTP и т. д.) Использование SQL-запросов. использование оператора "JOIN". Использование инструментов Business Intelligence (BI). BI-инструменты (QlikView, Tableau, Power BI). <i>Рекомендуемая литература 8.1-8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 3 Парсинг веб-страниц (веб-скрапинг). Мультипроцессинг.	Использование библиотеки BeautifulSoup, Использование библиотеки Scrapy. Использование модуля `multiprocessing`. Использование модуля `concurrent.futures`. <i>Рекомендуемая литература 8.1-8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 4 Получение данных с wiki / wikidata. Обход блокировок при парсинге сайта.	API для Википедии. API для Wikidata. Настраиваемые заголовки (Custom headers). Прокси-сервер (Proxy server). Задержка (Delay) <i>Рекомендуемая литература 8.1-8.5</i>	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6. Учебно-методическое обеспечение 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
--	--	--

Тема 1 Основные форматы данных. REST API. Запросы на динамический сайт.	REST API (Representational State Transfer) - стандарт Архитектуры Машинного Обучения для создания web-сервисов. REST API - Основные форматы данных и их применение. HTTP-методы для получения информации о ресурсе. Использование библиотек 'Requests' и 'Selenium'.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 2. Командные запросы: "Джек запросы". Протоколы (HTTP, FTP, SMTP и т. д.) для Curl запросов. Способы объединения (Merge) данных из разных источников.	- Протоколы для Curl запросов - Curl запросы для тестирования и проверки API, - Curl запросы для получения доступа к удаленному ресурсу и - Curl запросы для мониторинга работоспособности сайтов и веб-приложений. - Настройка Curl запросов под конкретные потребности - Использование SQL-запросов, использование оператора "JOIN". - Использование инструментов Business 'Intelligence' (BI) - 'QlikView', 'Tableau', 'Power BI. - Слияние данных из разных источников в единую таблицу. BI-системы - Объединение данные из разных баз данных, файлов или API. - Инструменты для интеграции данных - 'Talend', 'Pentaho', 'Informatica'	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 3 Парсинг веб-страниц (веб-скрапинг). Мультипроцессинг.	- автоматизация процесса парсинга веб-страниц, использование библиотек 'BeautifulSoup', 'Scrapy', 'multiprocessing', 'concurrent.futures'. - Использование нескольких процессов для выполнения различных задач одновременно на многоядерном или многопроцессорном компьютере. - Модуль 'multiprocessing'. - Модуль 'concurrent.futures'.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Тема 4 Получение данных с wiki / wikidata. Обход блокировок при парсинге сайта.	- API для Википедии. использование библиотеки 'wikipediaapi'. - API для Wikidata, использование библиотеки 'pywikibot'. - Методы обхода некоторых типов блокировок. - Настраиваемые заголовки (Custom headers), библиотека 'requests', чтобы запрос выглядел, как будто его отправил браузер. - Прокси-сервер (Proxy server) - перенаправление запросов с компьютера на сайт, для сокрытия ваш IP-адреса. - Задержка (Delay), для равномерного распределения загрузки на сайт, для избежания блокировки.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

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

Примерные вопросы к контрольной работе

1. Описать форматы данных - CSV, XML, JSON
2. Что такое REST API?
3. Как написать "Джак запросы" - через командную строку?
4. Как объединить данные из разных источников?
5. Какие библиотеки используются для парсинга веб-страниц?
6. Какие библиотеки используются для Мультипроцессинга?
7. Какие API можно использовать для получения данных из Википедии?

Примерные задания контрольной работы

1. Создать GET-запрос на динамический сайт с использованием библиотеки Selenium.
2. Создать обращение к своему бэкэнду с использованием библиотеки Requests.
3. Написать "Джак запросы" - через командную строку
4. Создать скрипт парсинга веб-страницы на Python для извлечения данных с сайта, с помощью BeautifulSoup.
5. Для двух задач, создать два процесса и запустить в каждую из задач в отдельном процессе.
6. Получить данные с Википедии используя её API
7. Создать запрос, чтобы он выглядел, как будто его отправил браузер.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента анализа данных и машинного обучения 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
ПКП- 3. Способность выполнять структурирование данных, выявлять признаки ценные для моделей машинного обучения	1. Демонстрирует знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.	<u>Знать:</u> Принципы процесса структурирования данных и формирования признаков для моделей машинного обучения. <u>Уметь:</u> Демонстрировать знание принципов процесса структурирования данных и формирования признаков для моделей машинного обучения.	Создать GET-запрос на динамический сайт с использованием библиотеки Selenium.
	2. Применяет инструментальные средств для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.	<u>Знать:</u> Как применять инструментальные средств для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения. <u>Уметь:</u> Применять инструментальные средств для преобразования данных и структурирования данных в рамках их предобработки для использования в моделях машинного обучения.	С помощью VI-системы объединить данные из двух разных файлов.

	3. Владеет практическим навыком инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.	<u>Знать:</u> Принципы инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи. <u>Уметь:</u> Использовать практические навыки инжиниринга признаков для моделей машинного обучения с учетом структуры исходных данных и предметной области прикладной задачи.	Написать GET-запрос с использованием библиотеки Requests
ПКП -1. Способность собирать наборы данных, в том числе больших данных, выполнять их подготовку для анализа в соответствии с решаемой прикладной задачей	1. Владеет навыками поиска внешних и внутренних источников данных для решения прикладной задачи.	<u>Знать:</u> Где искать внешние и внутренние источники данных для решения прикладной задачи. <u>Уметь:</u> Использовать навыки поиска внешних и внутренних источников данных для решения прикладной задачи.	Создать словарь `proxies`, содержащий URL-адреса прокси-сервера для протокола HTTPS.
	2. Использует инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников.	<u>Знать:</u> Инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников. <u>Уметь:</u> Использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников.	Написать SQL-запрос для объединения таблиц, с использованием оператора "JOIN".
	3. Владеет навыками обработки больших данных для решения прикладных задач машинного обучения.	<u>Знать:</u> Как использовать навыки обработки больших данных для решения прикладных задач машинного обучения. <u>Уметь:</u> Использовать навыки обработки больших данных для решения прикладных задач машинного обучения.	Реализовать мультипроцессинг, с использованием модуля `concurrent.futures`

Примерные вопросы для подготовки к экзамену

1. Описать форматы данных - CSV, XML, JSON
2. Что такое REST API?

3. Как написать "Джак запросы" - через командную строку?
4. Как объединить данные из разных источников?
5. Какие библиотеки используются для парсинга веб-страниц?
6. Какие библиотеки используются для Мультипроцессинга?
7. Какие API можно использовать для получения данных из Википедии?
8. В каких случаях рекомендуется использовать формат данных JSON, а в каких - XML?
9. Что такое формат данных CSV и в каких областях он наиболее широко применяется?
10. В чём преимущества использования формата данных YAML по сравнению с JSON и XML?
11. Какие основные типы данных могут быть хранены в формате данных INI и какие задачи он позволяет решать?
12. Какие типы запросов можно отправлять на REST API, и для чего они используются?
13. Какие из типов запросов, таких как GET, POST, PUT, PATCH и DELETE, используются для создания, получения, обновления и удаления данных на динамическом сайте?
14. Что такое REST API, какие принципы лежат в его основе, и как его можно использовать для взаимодействия с динамическими сайтами?
15. Какие инструменты могут использоваться для отправки запросов на REST API и обработки ответов, и как выбрать наиболее подходящий инструмент для конкретной задачи?
16. Что такое "Джак запросы" и как они используются для отправки командных запросов на сервере? Какие параметры могут быть включены в запросы?
17. Какие протоколы можно использовать для отправки Curl запросов, и что определяет выбор конкретного протокола?
18. Что такое HTTP-запросы, как они используются для обмена данными между клиентом и сервером, и какие типы запросов можно отправлять?

19. Какие основные протоколы используются для передачи файлов, например, для загрузки файлов на сервер или их скачивания? Как Curl запросы связаны с этими протоколами?
20. Как объединить данные из разных источников, таких как базы данных, CSV-файлы и API, чтобы получить единый набор данных?
21. Какие инструменты и технологии могут использоваться для объединения данных из разных источников, и как выбрать наиболее подходящий вариант в зависимости от конкретной задачи?
22. Какие трудности могут возникнуть при объединении данных из разных источников, и как их можно решить?
23. Какие методы объединения данных наиболее популярны в различных областях, например, в бизнесе, аналитике или научных исследованиях?
24. Что такое парсинг веб-страниц, и как он используется для извлечения данных из HTML и других форматов страниц?
25. Какие трудности могут возникнуть при парсинге веб-страниц, и как их можно решить?
26. Какой тип парсера лучше выбрать для конкретной задачи, и какие основные параметры нужно учитывать при выборе парсера?
27. Какие библиотеки и инструменты можно использовать для парсинга веб-страниц, и как выбрать наиболее эффективный и удобный вариант?
28. Каковы основные принципы работы веб-скрейпинга, и какие типы данных можно извлекать с веб-страниц?
29. Какие инструменты и библиотеки можно использовать для веб-скрапинга, и как выбрать наиболее подходящий вариант в зависимости от задачи?
30. Какие правила нужно соблюдать при проведении веб-скрапинга, и как избежать возможных проблем, связанных с авторским правом и другими ограничениями?
31. Как быстро и эффективно проводить веб-скрапинг с большого количества страниц и сайтов, и какие основные техники для этого существуют?

32. Что такое мультипроцессинг, как он работает и какие преимущества он предоставляет для параллельной обработки данных?
33. Какие языки программирования и инструменты поддерживают мультипроцессинг, и как выбрать самый подходящий вариант для конкретной задачи?
34. Что такое директивы OpenMP, как они работают и как использовать их для мультипроцессного программирования?
35. Какие трудности могут возникнуть при мультипроцессинге, и как их можно решить?
36. Какие типы данных предоставляются в Wiki / Wikidata, и как использовать API для получения данных?
37. Какие языки программирования и инструменты можно использовать для получения данных из Wiki / Wikidata, и как выбрать наиболее подходящий вариант в зависимости от задачи?
38. Какие трудности могут возникнуть при получении данных из Wiki / Wikidata, и как их можно решить?
39. Какие типы запросов поддерживаются в API Wikidata, и какие параметры может включать запрос для получения конкретных данных?
40. Почему сайты могут блокировать IP-адреса и как это может помешать парсингу сайта?
41. Какие техники и инструменты можно использовать для обхода блокировок при парсинге сайта, и как выбрать наиболее подходящий вариант в зависимости от задачи?
42. Какие меры предосторожности нужно принимать при обходе блокировок, чтобы избежать проблем с безопасностью и законности?
43. Как измерить отказоустойчивость парсера, работающего через обход блокировок?

Пример экзаменационного билета

- 1) Получить единый набор данных из двух разных источников, CSV и API **(20 баллов)**
- 2) Что такое REST API, какие принципы лежат в его основе, и как его можно использовать для взаимодействия с динамическими сайтами? **(20 баллов)**
- 3) В чём преимущества использования формата данных YAML по сравнению с JSON и XML? **(20 баллов)**

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

Основная литература:

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Юрайт, 2023. — 490 с. — (Высшее образование). — ЭБС Юрайт. — URL: <https://urait.ru/bcode/511020> (дата обращения: 19.01.2024). — Текст : электронный.
2. Мартишин, С. А. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala: учебное пособие / С. А. Мартишин, В. Л. Симонов, М. В. Храпченко. — Москва : ИНФРА-М, 2021. — 235 с. — ЭБС ZNANIUM. — URL: <http://znanium.com/catalog/product/1189321> (дата обращения: 19.01.2024). — Текст : электронный.

Дополнительная литература:

3. Колдаев, В. Д. Основы алгоритмизации и программирования : учебное пособие /под ред. проф. Л. Г. Гагариной. — Москва : Издательский Дом "ФОРУМ": ИНФРА-М, 2017. — 416 с. — ЭБС ZNANIUM. — URL: <http://znanium.com/catalog.php?bookinfo=902236> (дата обращения: 19.01.2024). — Текст : электронный.
4. Гуриков, С. Р. Основы алгоритмизации и программирования на Python : учебное пособие / С. Р. Гуриков. — Москва : ИНФРА-М, 2022. — 343 с. — (Высшее

образование: Бакалавриат). – ЭБС ZNANIUM. - URL: <https://znanium.com/catalog/product/1356003> (дата обращения: 19.01.2024). – Текст : электронный.

5. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения: Практическое пособие / С. Рашка; пер.с англ. А. В. Логунова. – Москва : ДМК Пресс, 2017. - 418 с. – ЭБС ZNANIUM. - URL: <http://znanium.com/go.php?id=1027758> (дата обращения: 19.01.2024). - 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. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
9. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. Национальная электронная библиотека <http://нэб.рф/>
12. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
13. СПАРК <https://spark-interfax.ru/>

10. Методические указания для обучающихся по освоению дисциплины

1. 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.
2. Необходимо конспектировать лекции, помечая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.
3. 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.
4. 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:

- Пакет офисных программ;
- Антивирус Kaspersky;

11.2. Современные профессиональные базы данных и информационные справочные системы:

- Информационно-правовая система «Гарант»;

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

11.3. Сертифицированные программные и аппаратные средства защиты информации: - не предусмотрены

11.4. Язык программирования Python 3.x в среде Windows.

11.5. Платформа для научных исследований, основанная на языке программирования Python, Anaconda, Jupyter.

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

Наличие аудитории, оснащенной компьютерной техникой и проектором, с возможностью подключения к сети «Интернет».