

**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**

УТВЕРЖДАЮ

Проректор по учебной
и методической работе

_____ Е.А. Каменева
25.04.2023 г.

Макрушин С.В., Блохин Н.В.

Технологии обработки больших данных

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

38.03.02 - Менеджмент,

ОП «Управление бизнесом»,

ОП "Управление бизнесом / Bachelor of Business Administration (BBA)",

ОП "Управление финансами / Bachelor of Business Administration in Finance",

ОП «Финансовый менеджмент»

Рекомендовано Ученым советом

*Факультета информационных технологий и анализа больших данных
(протокол №31 от 18.04.2023г.)*

Одобрено Советом учебно-научного

*Департамента анализа данных и машинного обучения
(протокол №2 от 29.03.2023г.)*

Москва 2023

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.....	5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	6
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	7
5.1. Содержание дисциплины	7
5.2. Учебно-тематический план.....	9
5.3. Содержание семинаров, практических занятий	12
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	14
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	14
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	16
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	17
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	25
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	26
10. Методические указания для обучающихся по освоению дисциплины.	27
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	28
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.	29

1. Наименование дисциплины

«Технологии обработки больших 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
ОП «Управление бизнесом», ОП «Управление бизнесом / Bachelor of Business Administration (BBA)», ОП «Управление финансами / Bachelor of Business Administration in Finance», ОП «Финансовый менеджмент»			
ПKN-5	Владение основами финансового учета и отчетности, а также принципами управленческого учета в целях использования данных учета для принятия управленческих решений	Применяет результаты анализа финансовой, бухгалтерской, управленческой отчетности при составлении стратегических финансовых, инвестиционных планов, отборе проектов и принятии управленческих решений.	Знать технологии обработки больших данных, применимые для анализа отчетности. Уметь применять технологии обработки больших данных, применимые для анализа отчетности.
		Анализирует и оценивает финансово-хозяйственное состояние организации и результаты деятельности их внутренних подразделений, формирует ключевые показатели эффективности для сбалансированного управления деятельностью организации.	Знать технологии обработки больших данных, применимые для комплексной обработки данных. Уметь применять технологии обработки больших данных, применимые для комплексной обработки данных.
ОП «Управление бизнесом», Профиль: Управление продуктом			
ПКП-3	Способность рассчитывать юнит-экономику продукта, разрабатывать требования к продукту и план развития продукта	Рассчитывает юнит-экономику продукта и разрабатывает требования к продукту.	Знать технологии для выполнения расчетов на больших объемах данных. Уметь выполнять расчеты на больших объемах данных.
		Управляет планом развития продукта.	Знать способы расчета сложных показателей и работы с

			различными форматами данных. Уметь выполнять расчеты сложных показателей и работать с различными форматами данных.
ОП «Управление бизнесом», Профиль: Менеджмент и управление бизнесом			
ПКП-3	Способность планировать, участвовать в процессах текущего и стратегического планирования и контроля в организации	Демонстрирует навыки использования в своей работе новых технологий для планирования и контроля в организации.	Знать современные технологии расчета показателей со сложными зависимостями. Уметь выполнять расчета показателей со сложными зависимостями с помощью современных технологий.
		Использует метрики результативности компании на каждом этапе жизненного цикла компании.	Знать современные технологии расчета экономических показателей на большом объеме данных. Уметь рассчитывать экономические показатели на большом объеме данных с помощью современных технологий.
ОП «Управление бизнесом / Bachelor of Business Administration (BBA)»			
Профиль: «Бизнес и предпринимательство / Business & Entrepreneurship»			
ПКП-3	Способность определять тенденции в развитии современной мировой и отечественной экономики, искать и находить возможности для создания бизнеса	Применяет теоретические знания и экономические законы для определения основных тенденций в развитии современной мировой экономики.	Знать технологии анализа больших данных, применимые для обработки экономических показателей. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей.
		Использует методики анализа последствий принимаемых управленческих решений в сфере международного бизнеса.	Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к глобальной экономике и международной торговле и финансам. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том

			числе относящихся к глобальной экономике и международной торговле и финансам.
Профиль «Управление маркетингом», «Управление маркетингом / Marketing Management»			
ПКП-2	Способность проводить маркетинговые исследования, анализировать конъюнктуру рынка и интерпретировать полученные результаты для принятия управленческих решений	Использует лучшие практики при планировании и проведении маркетинговых исследований.	Знать технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований.
		Демонстрирует навыки в получении маркетинговой информации с целью оценки конъюнктуры рынка.	Знать технологии анализа больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры. Уметь применять технологии анализа больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры.
ОП «Управление финансами / Bachelor of Business Administration in Finance», Профиль: «Управление финансами/BBA in Finance»			
ПКП-3	Способность разрабатывать стратегию ценообразования, осуществлять корректировку цены на основе исследования рыночной конъюнктуры	1.Разрабатывает стратегию ценообразования.	Знать технологии агрегации и анализа ценовой информации, представленной в больших массивах данных. Уметь использовать технологии агрегации и анализа ценовой информации, представленной в больших массивах данных.
		2.Осуществляет корректировку цены на основе исследований рыночной конъюнктуры.	Знать технологии построения прогнозных моделей на основе обработки больших массивов числовой информации. Уметь применять технологии построения прогнозных моде-

			лей на основе обработки больших массивов числовой информации.
ОП «Финансовый менеджмент», Профиль: Финансовый менеджмент			
ПКП-1	Способность оценивать тенденции и закономерности развития внешней и внутренней экономической среды, ее влияние на результаты хозяйственной деятельности организации в текущей и долгосрочной перспективе	Проводит необходимые для решения финансовых задач организации исследования внешней и внутренней среды с использованием современных информационных технологий.	Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к исследованиям внешней и внутренней среды организации. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к исследованиям внешней и внутренней среды организации.
		Оценивает и прогнозирует закономерности развития внешней и внутренней среды бизнеса.	Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к прогнозированию закономерностей развития внешней и внутренней среды бизнеса. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к прогнозированию закономерностей развития внешней и внутренней среды бизнеса.

3. Место дисциплины в структуре образовательных программ

Дисциплина «Технологии обработки больших данных» является дисциплиной Цикла профиля (элективный) по направлению подготовки 38.03.02 – Менеджмент, ОП «Управление бизнесом», ОП «Управление бизнесом / Bachelor of Business Administration (BBA)», ОП «Управление финансами / Bachelor of Business Administration in Finance», ОП «Финансовый менеджмент».

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

ОП «Управление бизнесом», ОП «Управление бизнесом / Bachelor of Business Administration (BBA)», ОП «Управление финансами / Bachelor of Business Administration in Finance», ОП «Финансовый менеджмент»

Очная форма обучения, 2021 / 2022 г.п. и т.д.

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 7 (в часах)
Общая трудоёмкость дисциплины	3/108	108
<i>Контактная работа- Аудиторные занятия</i>	<i>50/36</i>	<i>50/36</i>
Лекции	16/2	16/2
Семинары, практические занятия	34/34	34/34
<i>Самостоятельная работа</i>	<i>58/72</i>	<i>58/72</i>
Вид текущего контроля	Зачет	Зачет
Вид промежуточной аттестации	Домашнее творческое задание	Домашнее творческое задание

ОП «Управление бизнесом»

очно-заочная форма обучения, 2021 / 2022 г.п. и т.д.

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 8 (в часах)
Общая трудоёмкость дисциплины	3/108	108
<i>Контактная работа- Аудиторные занятия</i>	<i>32/18</i>	<i>32/18</i>
Лекции	16/2	16/2
Семинары, практические занятия	16/16	16/16
<i>Самостоятельная работа</i>	<i>76/90</i>	<i>76/90</i>
Вид текущего контроля	Зачет	Зачет
Вид промежуточной аттестации	Домашнее творческое задание	Домашнее творческое задание

Институт онлайн-образования
ОП «Финансовый менеджмент»

Очно-заочная форма обучения, 2021, 2022 г.п. и т.д.

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 8 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа-Аудиторные занятия	32	32
Лекции	8	8
Семинары, практические занятия	24	24
Самостоятельная работа	76	76
Вид текущего контроля	Зачет	Зачет
Вид промежуточной аттестации	Домашнее творческое задание	Домашнее творческое задание

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 numpy.

Тема 6. Взаимодействие с базой данных в приложениях обработки данных.

В рамках темы рассматривается взаимодействие из Python с базой данных на примере API SQLite. Базовые возможности работы с транзакциями.

5.2. Учебно-тематический план

ОП «Управление бизнесом», ОП «Управление бизнесом / Bachelor of Business Administration (BBA)», ОП «Управление финансами / Bachelor of Business Administration in Finance», ОП «Финансовый менеджмент»

очная форма обучения, 2021 / 2022 г.п. и т.д.

№ 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, практические занятия		
1	Библиотека NumPy и Pandas	22/22	12/10	4/2	8/8	10/12	Участие в решении задач на практических занятиях. Обсуждения по результатам самостоятельной работы
2	Использование различных форматов файлов в задачах обработки данных.	20/16	10/4	4/0	6/4	10/12	
3	Взаимодействие с табличными данными в приложениях обработки данных.	14/18	6/6	2/0	4/6	8/12	
4	Визуализация данных	16/18	8/6	2/0	6/4	8/12	
5	Работа со строками в приложениях обработки данных	18/16	6/4	2/0	4/4	12/12	Участие в решении задач на практических занятиях. Обсуждения по результатам самостоятельной работы
6	Взаимодействие с базой данных в приложениях обработки данных.	18/16	8/4	2/0	6/8	10/12	
	В целом по дисциплине	108	50/36	16/2	34/34	58/72	Согласно учебному плану: домашнее творческое задание
	Итого в %		46/33	32/2	68/31	54/67	

ОП «Управление бизнесом»

очно-заочная форма обучения, 2021 / 2022 г.п. и т.д.

№ 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, практические занятия		
1	Библиотека NumPy и Pandas	22/22	8/6	4/2	4/4	14/16	Участие в ре- шении задач на практиче- ских занятиях. Обсуждения по результа- там самосто- ятельной ра- боты Участие в ре- шении задач на практиче- ских занятиях. Обсуждения по результа- там самосто- ятельной ра- боты
2	Использование различных форма- тов файлов в зада- чах обработки данных.	16/18	4/2	2/0	2/2	12/16	
3	Взаимодействие с табличными дан- ными в приложе- ниях обработки данных.	16/16	4/2	2/0	2/2	12/14	
4	Визуализация данных	16/16	4/2	2/0	2/2	12/14	
5	Работа со стро- ками в приложе- ниях обработки данных	16/16	4/2	2/0	2/2	12/14	
6	Взаимодействие с базой данных в приложениях об- работки данных.	22/20	8/4	4/0	4/4	14/16	
	В целом по дисци- плине	108	32/18	16/2	16/16	76/90	Согласно учебному плану: домаш- нее творче- ское задание
	Итого в %		30/17	50/11	50/89	70/83	

ОП «Финансовый менеджмент»

очно-заочная форма обучения (ИОО), 2021, 2022 г.п. и т.д.

№ 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, практические занятия		
1	Библиотека NumPy и Pandas	20	6	2	4	14	Участие в решении задач на практических занятиях. Обсуждения по результатам самостоятельной работы
2	Использование различных форматов файлов в задачах обработки данных	17	5	1	4	12	
3	Взаимодействие с табличными данными в приложениях обработки данных	17	5	1	4	12	
4	Визуализация данных	17	5	1	4	12	
5	Работа со строками в приложениях обработки данных	17	5	1	4	12	
6	Взаимодействие с базой данных в приложениях обработки данных.	20	6	2	4	14	
	В целом по дисциплине	108	32	8	24	76	Согласно учебному плану: домашнее творческое задание
	Итого в %		30	25	75	70	

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	• Технологический стек Python для обработки и анализа данных • Возможности Python как glue language • Организация массивов в NumPy: хранение данных, создание массивов • Принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. • 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. 8[1], 9[9], 9[10]	Интерактивная форма, работа на компьютере
Использование различных форматов файлов в задачах обработки данных	• Формат файлов Pickle, представление данных в этом формате и взаимодействие с ним в Python. • Формат файлов JSON, представление данных в этом формате и взаимодействие с ним в Python. • Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM • Работа с XML с помощью библиотеки BeautifulSoup. 8[1], 8[2], 9[3], 9[4]	Интерактивная форма, работа на компьютере
Взаимодействие с табличными данными в приложениях обработки данных.	• Взаимодействие с Excel из Python с помощью библиотеки XLWings. • Формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python 8[1], 8[2]	Интерактивная форма, работа на компьютере
Визуализация данных	• Построение визуализаций с помощью библиотеки matplotlib • Построение визуализаций с помощью библиотеки pandas • Построение визуализаций с помощью библиотеки seaborn 8[1], 9[13], 9[15], 9[16]	Интерактивная форма, работа на компьютере
Работа со строками в приложениях обработки данных	• Основы работы с регулярными выражениями: базовый синтаксис, примеры. • Модуль re в Python. 8[1], 8[2], 9[4]	Интерактивная форма, работа на компьютере
Взаимодействие с ба-	• Взаимодействие из Python с базой данных с помощью API SQLite.	Интерактивная форма, работа на

зой данных в приложениях обработки данных	8[1], 8[2]	компьютере
Профилирование процессов обработки данных, библиотека Numba.	• профилирование реализации алгоритмов на Python • принципы решения задачи оптимизации производительности алгоритма • Библиотека Numba: принципы работы, базовые примеры использования. 8[1], 8[2], 9[1], 9[2], 9[3]	Интерактивная форма, работа на компьютере
Параллельная обработка данных, введение в Dask	• специфика современного аппаратного обеспечения для обработки больших данных и проблема масштабируемости параллельных вычислений. • Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения. • Проблема Global Interpreter Lock в Python и способы обхода ее ограничений. • Модуль Python multiprocessing – назначение и основные возможности, API multiprocessing.Pool. • Подход к обработке данных с помощью библиотеки Dask. • Структура данных Dask.Array – принцип работы, API, примеры использования. • Структура данных Dask.DataFrame – принцип работы, API, примеры использования 8[1], 8[2], 9[5], 9[7], 9[8], 9[10], 9[11]	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	• Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры. • Маскирование и прихотливое индексирование в NumPy. • Операция GroupBy в Pandas DataFrame и	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.

	реализация в ней подхода «разбиение, применение и объединение».	
Использование различных форматов файлов в задачах обработки данных	• Формат файлов NPY, представление данных в этом формате и взаимодействие с ним в Python. • Формат файлов HDF, представление данных в этом формате и взаимодействие с ним в Python.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Взаимодействие с табличными данными в приложениях обработки данных.	• Продвинутое взаимодействие с Excel из Python с помощью библиотеки XLWings.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Визуализация данных	• Построение трехмерных графиков Продвинутое взаимодействие с цветовыми картами	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Работа со строками в приложениях обработки данных	• Использование хэширования при работе со строками. • Строки в библиотеке numpy.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Взаимодействие с базой данных в приложениях обработки данных	• Базовые возможности работы с транзакциями с помощью API SQLite.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Профилирование процессов обработки данных, библиотека Numba.	• Векторизация в numpy: ключевые параметры функции, примеры применения • Использование обобщенной сигнатуры функции в numpy и numba. •	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Параллельная обработка данных, введение в Dask	• Модели параллельного программирования и их сочетаемость с архитектурами параллельных вычислительных систем. • Специфика различия между потоками и процессами. • Организация вычислений с помощью Map / Filter / Reduce: общий принцип и специфика параллельной реализации обработки данных с помощью Dask.Bag. • Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.

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

Примерная тематика домашнего творческого задания

1. Прогнозная аналитика и моделирование объемов продаж акций
2. Визуализация аналитических данных в области макроэкономики
3. Визуализация аналитических данных Московской биржи
4. Использование технологии больших данных для анализа портфельных рисков
5. Использование параллельных вычислений реализации численных методов решения математических задач
6. Анализ и сравнение различных фреймворков для визуализации данных
7. Применение распределенных вычислений и экосистемы Hadoop для решения задачи анализа данных
8. Анализ больших данных для построения прогнозов на рынке ценных бумаг
9. Использование больших данных для оценки кредитоспособности контрагентов на основе анализа текстов новостей
10. Проведение анализа собранных из внешних источников данных

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
ОП «Управление бизнесом», ОП «Управление бизнесом / Bachelor of Business Administration (BBA)», ОП «Управление финансами / Bachelor of Business Administration in Finance», ОП «Финансовый менеджмент»			
ПКН-5. Владение основами финансового учета и отчетности, а также принципами управленческого учета в целях использования данных учета для принятия управленческих решений	Применяет результаты анализа финансовой, бухгалтерской, управленческой отчетности при составлении стратегических финансовых, инвестиционных планов, отборе проектов и принятии управленческих решений.	Знать технологии обработки больших данных, применимые для анализа отчетности. Уметь применять технологии обработки больших данных, применимые для анализа отчетности.	Предложите технологический стек для анализа первичной отчетности объемом 100Мб, 1Гб, 10Гб. Реализуйте алгоритм агрегации показателей для набора данных объемом 1Гб.
	Анализирует и оценивает финансово-хозяйственное состояние организации и результаты деятельности их внутренних подразделений, формирует ключевые показатели эффективности для сбалансиро-	Знать технологии обработки больших данных, применимые для комплексной обработки данных. Уметь применять технологии обработки больших данных, применимые	Предложите технологический стек для многокритериальной группировки данных первичной отчетности объемом 100Мб, 1Гб, 10Гб. Реализуйте алгоритм многокритериальной группировки данных пер-

	рованного управле- ния деятельностью организации.	для комплексной об- работки данных.	вичной отчетности объе- мом 1Гб.
ОП «Управление бизнесом»			
Профиль: Управление продуктом			
ПКП-3. Спо- собность рас- считывать юнит-эконо- мику продукта, разрабатывать требования к продукту и план развития продукта	1. Рассчитывает юнит-экономику продукта и разраба- тывает требования к продукту.	Знать современные технологии расчета показателей со сложными зависи- мостями. Уметь выполнять расчета показателей со сложными зави- симостями с помо- щью современных технологий.	Предложите технологи- ческий стек для расчета показателей юнит-эконо- мики, объем которых со- ставляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм рас- чета показателей юнит- экономики, объем данных для которых составляет 100Мб.
	2. Управляет планом развития продукта.	Знать современные технологии расчета экономических по- казателей на боль- шом объеме данных. Уметь рассчитывать экономические по- казателей на боль- шом объеме данных с помощью совре- менных технологий.	Предложите технологи- ческий стек для расчета экономических показате- лей, объем которых со- ставляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм рас- чета экономических пока- зателей (например, при- были по ЦФО), объем данных для которых со- ставляет 100Мб.
Профиль: Менеджмент и управление бизнесом			
ПКП-3. Спо- собность пла- нировать, участвовать в процессах теку- щего и страте- гического пла- нирования и контроля в ор- ганизации	1. Демонстрирует навыки использова- ния в своей работе новых технологий для планирования и контроля в организа- ции.	Знать технологии анализа больших данных, примени- мые для обработки экономических по- казателей. Уметь применять технологии анализа больших данных, применимые для об- работки экономиче- ских показателей.	Предложите технологи- ческий стек для расчета и обработки экономиче- ских показателей, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм рас- чета экономических пока- зателей и планирования (например, производ- ственного плана), объем данных для которых со- ставляет 100Мб.
	2. Использует мет- рики результативно- сти компании на каж- дом этапе жизнен-	Знать технологии анализа больших данных, примени- мые для обработки	Предложите технологи- ческий стек для расчета экономических показате- лей международной эконо-

	ного цикла компании.	экономических показателей, в том числе относящихся к глобальной экономике и международной торговле и финансам. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к глобальной экономике и международной торговле и финансам.	номики, торговли и финансов объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм расчета экономических показателей международной экономики, торговли и финансов (например, транзакций на крупнейших глобальных биржевых площадках), объем данных для которых составляет 100Мб.
ОП «Управление бизнесом / Bachelor of Business Administration (BBA)»			
Профиль: «Бизнес и предпринимательство / Business & Entrepreneurship»			
ПКП-3. Способность определять тенденции в развитии современной мировой и отечественной экономики, искать и находить возможности для создания бизнеса	1.Применяет теоретические знания и экономические законы для определения основных тенденций в развитии современной мировой экономики.	Знать технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований.	Предложите технологический стек для расчета экономических и маркетинговых показателей, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм расчета маркетинговых показателей (например, данных о показах баннерной рекламы), объем данных для которых составляет 100Мб.
	2.Использует методики анализа последствий принимаемых управленческих решений в сфере международного бизнеса.	Знать технологии анализа больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры. Уметь применять технологии анализа	Предложите технологический стек для расчета и анализа экономических и маркетинговых показателей, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм анализа маркетинговых по-

		больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры.	казателей (например, прогноза эффективности баннерной рекламы), объем данных для которых составляет 100Мб.
Профили: «Управление маркетингом», «Управление маркетингом / Marketing Management»			
ПКП-2. Способность проводить маркетинговые исследования, анализировать конъюнктуру рынка и интерпретировать полученные результаты для принятия управленческих решений	1.Использует лучшие практики при планировании и проведении маркетинговых исследований.	Знать технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей и показателей маркетинговых исследований.	Предложите технологический стек для расчета экономических и маркетинговых показателей, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм расчета маркетинговых показателей (например, обработки результатов А/В тестирования), объем данных для которых составляет 100Мб.
	2.Демонстрирует навыки в получении маркетинговой информации с целью оценки конъюнктуры рынка.	Знать технологии анализа больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры. Уметь применять технологии анализа больших данных, применимые для обработки маркетинговых показателей и оценки маркетинговой конъюнктуры.	Предложите технологический стек для обработки экономических и маркетинговых показателей, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм обработки маркетинговых показателей (например, прибыли по ЦФО), объем данных для которых составляет 100Мб.
ОП "Управление финансами / Bachelor of Business Administration in Finance", Профиль: "Управление финансами/BBA in Finance"			
ПКП-3. Способность разрабатывать стратегию ценооб-	1.Разрабатывает стратегию ценообразования.	Знать технологии агрегации и анализа ценовой информации, представленной	Для таблицы большого объема с ценами для разных SCU в разных регионах и по разным датам рассчитать с помощью

разования, осуществлять корректировку цены на основе исследования рыночной конъюнктуры		в больших массивах данных. Уметь использовать технологии агрегации и анализа ценовой информации, представленной в больших массивах данных.	технологий обработки больших данных рассчитывать средние цены по региону и укрупненному периоду (месяцу).
	2.Осуществляет корректировку цены на основе исследований рыночной конъюнктуры	Знать технологии построения прогнозных моделей на основе обработки больших массивов числовой информации. Уметь применять технологии построения прогнозных моделей на основе обработки больших массивов числовой информации.	Для таблицы большого объема с ценами для разных SCU в разных регионах и по разным датам рассчитать с помощью технологий обработки больших данных рассчитать на основе экстраполяции прогноз средних цен по региону и укрупненному периоду (месяцу).
ОП «Финансовый менеджмент», Профиль: Финансовый менеджмент			
ПКП-1. Способность оценивать тенденции и закономерности развития внешней и внутренней экономической среды, ее влияние на результаты хозяйственной деятельности организации в текущей и долгосрочной перспективе	1. Проводит необходимые для решения финансовых задач организации исследования внешней и внутренней среды с использованием современных информационных технологий.	Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к исследованиям внешней и внутренней среды организации. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к исследованиям внешней и внутренней среды организации.	Предложите технологический стек для обработки экономических показателей, характеризующих внешнюю и внутреннюю среду организации, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм обработки экономических показателей, характеризующих внешнюю и внутреннюю среду организации (например, прибыли по ЦФО организации), объем данных для которых составляет 100Мб.

	2. Оценивает и прогнозирует закономерности развития внешней и внутренней среды бизнеса.	Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к прогнозированию закономерностей развития внешней и внутренней среды бизнеса. Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к прогнозированию закономерностей развития внешней и внутренней среды бизнеса.	Предложите технологический стек для обработки экономических показателей, в том числе прогнозирования внешней и внутренней среды организации, объем которых составляет 10Мб, 100Мб, 1Гб. Реализуйте алгоритм прогнозирования экономических показателей, характеризующих внешнюю и внутреннюю среду организации (например, прогноз прибыли по ЦФО организации), объем данных для которых составляет 100Мб.
--	--	---	--

Примерные вопросы для подготовки к зачету

1. Большие данные – определение и причины возникновения задач обработки больших данных
2. Специфика современного аппаратного обеспечения для обработки больших данных и проблема масштабируемости параллельных вычислений
3. Выбор типичных средств обработки данных, адекватных различным объемам данных; принцип обработки данных на базе операций map / filter / reduce
4. Многопроцессорные архитектуры с общей и разделяемой памятью – специфика и сравнение
5. Подходы к декомпозиции крупных вычислительных задач на подзадачи для параллельного исполнения

6. Модели параллельного программирования и их сочетаемость с архитектурами параллельных вычислительных систем
7. Профилирование реализации алгоритмов на Python, принципы решения задачи оптимизации производительности алгоритма
8. Проблема Global Interpreter Lock в Python и способы обхода ее ограничений
9. Технологический стек Python для обработки и анализа данных, Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python
10. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
11. Универсальные функции и применение функций по осям в NumPy
12. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
13. Маскирование и прихотливое индексирование в NumPy
14. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции
15. Numba: принципы работы, базовые примеры использования
16. Организация Pandas DataFrame и организация индексации для DataFrame и Series
17. Применение универсальных функций и работа с пустыми значениями в Pandas
18. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры
19. Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение»
20. Специфика текстовых и бинарных файлов, форматы файлов CSV и Pickle, представление данных в этих форматах и взаимодействие с ними в Python

21. Задача сериализации и десериализации, описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python
22. Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup
23. Форматы файлов NPY и HDF общая характеристика, пример взаимодействие с данными этих форматов в Python
24. Взаимодействие из Python с базой данных на примере API SQLite, базовые возможности работы с транзакциями
25. Взаимодействие с Excel из Python с помощью XLWings: принципы работы и примеры использования
26. Основы работы с регулярными выражениями: базовый синтаксис, примеры использования модуля re в Python
27. Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python
28. Различия между потоками и процессами, различие между различными планировщиками в Dask
29. Граф зависимостей задач – суть структуры данных, ее построение и использование в Dask
30. Три ключевых структуры данных Dask: их специфика и принцип выбора структуры данных при решении задач
31. Dask.Array – структура данных, специфика реализации и применения, процедура создания
32. Dask.Array – поддерживаемые операции и отличия от NumPy ndarray
33. Распараллеливание алгоритмов с помощью dask.delayed – принцип и примеры использования
34. Дополнительные параметры декоратора dask.delayed – назначение и примеры использования

35. Использование `dask.delayed` для объектов и операции над объектами `dask.delayed`, включая ограничения их использования
36. `Dask.DataFrame` - структура данных, специфика реализации и применения, процедура создания `Dask.DataFrame`
37. Ограничения использования `Dask.DataFrame` и операции мэппинга в `Dask.DataFrame`
38. Поддержка `Dask.DataFrame` операций работающих со скользящим окном
39. Совместное использование промежуточных результатов в Dask: принцип работы и примеры использования
40. `Dask.Bag` - структура данных, специфика реализации и применения, процедура создания `Dask.Bag`
41. Организация вычислений с помощью Map / Filter / Reduce : общий принцип и специфика параллельной реализации обработки данных в `Dask.Bag`
42. Понятие признака в анализе данных и типы признаков
43. Понятие разведочного анализа данных, основные задачи и типовые визуализации для решения этих задач

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.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 15.06.2023). – Текст : электронный.

Дополнительная литература:

2. 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: 15.06.2023). – 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.znaniy.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. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>

11. Национальная электронная библиотека <http://нэб.рф/>

12. Pyru 1.0.9 [Электронный ресурс]: сайт. – Режим доступа:
<https://pypi.python.org/pypi/pyru>

13. Python Data Analysis Library [Электронный ресурс]: сайт. – Режим доступа: <http://pandas.pydata.org/>

14. Python Documentation [Электронный ресурс]: сайт. – Режим доступа: <http://python.org/doc/>

15. Python Standard Library [Электронный ресурс]: сайт. – Режим доступа: <https://docs.python.org/2/library/>

16. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. – Режим доступа: <http://scikit-learn.org>

17. Официальный сайт продукта <https://www.python.org/>

18. Каталог курсов Интернет Университета Информационных Технологий <http://www.intuit.ru/>

19. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>

20. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>

21. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>

22. Dask User Guide <https://docs.dask.org/en/latest/>

23. Dask User Guide <https://docs.dask.org/en/latest/>

24. Matplotlib User Guide // <https://matplotlib.org/stable/users/index.html>

25. Seaborn User Guide // <https://seaborn.pydata.org/tutorial.html>

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.