

**«Финансовый университет при Правительстве Российской Федерации»
(Финуниверситет)**

Калужский филиал Финуниверситета

Кафедра «Бизнес – информатика и высшая математика»

«УТВЕРЖДАЮ»

**Директор Калужского филиала
Финуниверситета**



В.А. Матчинов

«30» июня 2022 г.

Пономарев С.В.

ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

38.03.02 «Менеджмент»

Образовательная программа «Управление бизнесом»

Очная форма обучения

*Рекомендовано Ученым советом Калужского филиала Финуниверситета
(протокол №56 от 30.06. 2022 г.)*

**Одобрено кафедрой «Бизнес – информатика и высшая математика»
Калужского филиала Финуниверситета
(протокол № 12 от 28 июня 2022 г.)**


КАЛУГА 2022

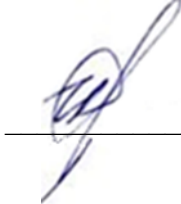
Рабочая программа предназначена для преподавания дисциплины «Технологии обработки больших данных» студентам, обучающимся по направлению подготовки 38.03.02 «Менеджмент», образовательная программа «Управление бизнесом» по очной форме обучения.

В рабочей программе излагаются планируемые результаты освоения дисциплины, содержание дисциплины, тематика и содержание семинаров и практических занятий, технологии их проведения. В рабочей программе дисциплины приводится перечень учебно-методического обеспечения для самостоятельной работы обучающихся, фонд оценочных средств для проведения промежуточной аттестации обучающихся, перечень основной и дополнительной литературы, а также ресурсов информационно-телекоммуникационной сети Интернет.

СОГЛАСОВАНО:

Заместитель директора
по учебно-методической работе  /Орловцева О.М./
«28» июня 2022 г.

Начальник учебно-методического отдела  /Толстикова В.С./
«28» июня 2022 г.

Заведующий кафедрой
«Бизнес-информатика и высшая математика»  /Дробышева И.В./
«28» июня 2022 г.

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

Б.1.2.2.2.4.2. «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
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.
		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.
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.	<p>Знать современные технологии расчета экономических показателей на большом объеме данных.</p> <p>Уметь рассчитывать экономические показатели на большом объеме данных с помощью современных технологий.</p>
--	--	---	---

3. Место дисциплины в структуре образовательных программ

Дисциплина «Технологии обработки больших данных» является дисциплиной модуля «IT Skills» по направлению подготовки 38.03.02 – Менеджмент, ОП «Управление бизнесом».

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

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 7 (в часах)
Общая трудоёмкость дисциплины	3/108	108
Контактная работа- Аудиторные занятия	36	36
Лекции	2	2
Семинары, практические занятия	34	34
Самостоятельная работа	72	72
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 дан-

ных на примере API SQLite. Базовые возможности работы с транзакциями.

5.2. Учебно-тематический план

Таблица 3.

№ 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	10	2	8	12	Участие в решении задач на практических занятиях. Обсуждения по результатам самостоятельной работы
2	Использование различных форматов файлов в задачах обработки данных.	16	4	-	4	12	
3	Взаимодействие с табличными данными в приложениях обработки данных.	18	6	-	6	12	
4	Визуализация данных	18	6	-	4	12	
5	Работа со строками в приложениях обработки данных	16	4	-	4	12	
6	Взаимодействие с базой данных в приложениях обработки данных.	16	4	-	8	12	
	В целом по дисциплине	108	36	2	34	72	Согласно учебному плану: домашнее творческое задание
	Итого в %		33	2	31	67	

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"> • 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. <p>8[1], 9[9], 9[10]</p>	Интерактивная форма, работа на компьютере
Использование различных форматов файлов в задачах обработки данных	<ul style="list-style-type: none"> • 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. <p>8[1], 8[2], 9[3], 9[4]</p>	Интерактивная форма, работа на компьютере
Взаимодействие с табличными данными в приложениях обработки данных.	<ul style="list-style-type: none"> • Взаимодействие с Excel из Python с помощью библиотеки XLWings. • Формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python <p>8[1], 8[2]</p>	Интерактивная форма, работа на компьютере
Визуализация данных	<ul style="list-style-type: none"> • Построение визуализаций с помощью библиотеки matplotlib • Построение визуализаций с помощью библиотеки pandas • Построение визуализаций с помощью библиотеки seaborn <p>8[1], 9[13], 9[15], 9[16]</p>	Интерактивная форма, работа на компьютере
Работа со строками в приложениях обработки данных	<ul style="list-style-type: none"> • Основы работы с регулярными выражениями: базовый синтаксис, примеры. • Модуль re в Python. <p>8[1], 8[2], 9[4]</p>	Интерактивная форма, работа на компьютере
Взаимодействие с ба-	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • Продвинутое взаимодействие с Excel из Python с помощью библиотеки XLWings. 	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Визуализация данных	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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 кафедры.

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
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 управления деятельностью организации.	для комплексной обработки данных.	вичной отчетности объемом 1Гб.
ПКП-3. Способность планировать, участвовать в процессах текущего и стратегического планирования и контроля в организации	1. Демонстрирует навыки использования в своей работе новых технологий для планирования и контроля в организации.	<p>Знать технологии анализа больших данных, применимые для обработки экономических показателей.</p> <p>Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей.</p>	<p>Предложите технологический стек для расчета и обработки экономических показателей, объем которых составляет 10Мб, 100Мб, 1Гб.</p> <p>Реализуйте алгоритм расчета экономических показателей и планирования (например, производственного плана), объем данных для которых составляет 100Мб.</p>
	2. Использует метрики результативности компании на каждом этапе жизненного цикла компании.	<p>Знать технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к глобальной экономике и международной торговле и финансам.</p> <p>Уметь применять технологии анализа больших данных, применимые для обработки экономических показателей, в том числе относящихся к глобальной экономике и международной торговле и финансам.</p>	<p>Предложите технологический стек для расчета экономических показателей международной экономики, торговли и финансов объем которых составляет 10Мб, 100Мб, 1Гб.</p> <p>Реализуйте алгоритм расчета экономических показателей международной экономики, торговли и финансов (например, транзакций на крупнейших глобальных биржевых площадках), объем данных для которых составляет 100Мб.</p>

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 при-

- менения, 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 DaskBag
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: 15.06.2022). – 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: 15.06.2022). – 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. 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 до-
ступа: <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 Техно-
логий <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

Обучающимся в рамках самостоятельной работы следует использовать Методические рекомендации по планированию и организации внеаудиторной самостоятельной работы студентов по образовательным программам бакалавриата и магистратуры в Финансовом университете, утвержденные Приказом ректора №1040/о от 11.05.2021 г.

Самостоятельная работа студентов проходит внеаудиторно. Организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В данном плане указана тематика лекций, семинаров, вопросы и задания для самостоятельного изучения. Во время лекций необходимо конспектировать содержание лекции. После лекции необходимо отредактировать записи, оформить конспект, дополняя его содержание дополнительной информацией. При оформлении конспекта целесообразно выделять названия тем и формулировки вопросов, основные определения, примеры.

При подготовке к семинару необходимо изучить вопросы семинара, соответствующий теоретический материал, делая для себя необходимые записи в рабочей тетради. После занятий необходимо просмотреть записанные решения и восстановить в решениях имеющиеся пробелы.

При затруднении в решении практических вопросов (задач), можно обратиться за консультацией (помощью) к преподавателю. Семинары проходят, как правило, в интерактивной форме и преподаватель учитывает активность обучающихся, направленную на решение предложенных вопросов (вариантов задач), а также вариантов ответов на решаемые вопросы (проблемы).

Не следует бояться дать неверный ответ или допустить иную ошибку: исправление и анализ ошибок в режиме общения с преподавателем и сокурсниками в ходе семинара способствует более глубокому освоению учебного материала и предупреждает возникновение ошибок в дальнейшем. Домашние задания (подготовку к занятиям) следует осуществлять регулярно. Если то или иное задание, при подготовке к семинару вызвало затруднение, необходимо обратиться к преподавателю за консультацией. Регулярность в выполнении домашних заданий (подготовке к занятиям) - важный фактор качественного освоения дисциплины.

Методические рекомендации по обучению лиц с ограниченными возможностями здоровья

Профессорско-преподавательский состав знакомится с психолого-физиологическими особенностями обучающихся инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ). При необходимости осуществляется дополнительная поддержка преподавания тьюторами, психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. № АК-44/05вн) в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного

психологического климата в студенческой группе. Подбор и разработка учебных материалов производится с учётом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем. Освоение дисциплины лицами с ОВЗ осуществляется с использованием средств обучения общего и специального назначения (персонального и коллективного использования). Материально-техническое обеспечение предусматривает приспособление аудиторий к нуждам лиц с ОВЗ. Форма проведения аттестации для студентов-инвалидов устанавливается с учётом индивидуальных психофизических особенностей.

Для студентов с ОВЗ предусматривается доступная форма предоставления заданий оценочных средств, а именно:

- в печатной или электронной форме (для лиц с нарушениями опорно-двигательного аппарата);
 - в печатной форме или электронной форме с увеличенным шрифтом и контрастностью (для лиц с нарушениями слуха, речи, зрения);
 - методом чтения ассистентом задания вслух (для лиц с нарушениями зрения).
- Студентам с инвалидностью увеличивается время на подготовку ответов на контрольные вопросы.

Для таких студентов предусматривается доступная форма предоставления ответов на задания, а именно:

- письменно на бумаге или набором ответов на компьютере (для лиц с нарушениями слуха, речи);
- выбором ответа из возможных вариантов с использованием услуг ассистента (для лиц с нарушениями опорно-двигательного аппарата);
- устно (для лиц с нарушениями зрения, опорно-двигательного аппарата). При необходимости для обучающихся с инвалидностью процедура оценивания результатов обучения может проводиться в несколько этапов.

Методические рекомендации по выполнению ДТЗ

Методические рекомендации по выполнению ДТЗ предусмотрены в «Методических рекомендациях по подготовке написанию и оформлению ДТЗ», разрабатываемой преподавателем кафедры на учебный год, в котором реализуется учебная дисциплина

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

11.1 Комплект лицензионного программного обеспечения

1. Антивирусная защита ESET NOD32
2. Windows, Microsoft Office
3. Astra Linux

11.2 Современные профессиональные базы данных и информационные справочные системы

1. Информационно-правовая система «Консультант Плюс»;
2. Информационно-правовая система «Гарант»;

11.3 Сертифицированные программные и аппаратные средства защиты информации

Сертифицированные программные и аппаратные средства защиты информации не предусмотрены

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

- аудитории для проведения лекционных и семинарских занятий, оборудованные видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения, экраном, доской меловой/интерактивной;
- библиотеку, имеющую рабочие места для студентов, оснащенные компьютерами с доступом к базам данных и сети Интернет
- компьютерные классы с набором лицензионного базового программного обеспечения для проведения практических занятий и выходом в глобальную сеть Internet;

Образовательные технологии

В процессе изучения дисциплины «Технологии обработки больших данных» предполагается:

- сопровождение курса лекций наглядной презентацией, включающей практические примеры, схемы, графики, табличный материал;
- рассмотрение на семинарских занятиях интерактивных ситуационных задач по проблематике дисциплины;
 - деловые игры;
 - разбор конкретных ситуаций, коллективное обсуждение проблем российской и зарубежной практики по изучаемым темам;
- виртуальное общение в течение срока изучения курса в целях обеспечения лекций и практических занятий необходимым материалом и также контроля самостоятельной работы студентов.